

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Boris Savić

**Tvorba jezika japonskih svečnikov in
uporaba NLP algoritma Word2Vec za
napovedovanje trendov gibanja
vrednosti delnic**

MAGISTRSKO DELO
ŠTUDIJSKI PROGRAM DRUGE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: doc. dr. Dejan Lavbič

Ljubljana, 2016

Rezultati magistrskega dela so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavljanje ali izkoriščanje rezultatov magistrskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorja.

Za mentorstvo in strokovno pomoč se zahvaljujem mentorju doc. dr. Dejanu Lavbiču. Zahvala gre tudi podjetju XLAB d.o.o. za finančno podporo skozi vsa leta študija.

Staršem.

Kazalo

Povzetek

Abstract

| | | |
|----------|---|-----------|
| 1 | Uvod | 1 |
| 2 | Podatki | 5 |
| 2.1 | Vir podatkov | 6 |
| 2.2 | Coca-Cola (NYSE:KO) | 7 |
| 2.3 | Microsoft (NASDAQ:MSFT) | 8 |
| 2.4 | Apple (NASDAQ:AAPL) | 9 |
| 2.5 | Russell Top 50 | 11 |
| 3 | Opis napovednega modela | 13 |
| 3.1 | Samodejna zaznava japonskih svečnikov | 15 |
| 3.2 | Tvorba OHLC jezika | 21 |
| 3.3 | Obdelava z NLP algoritmom Word2Vec | 23 |
| 3.4 | Napovedovanje | 27 |
| 3.5 | Trgovalna strategija | 30 |
| 4 | Metrike uspešnosti | 35 |
| 4.1 | Napovedni model | 36 |
| 4.2 | Trgovalna strategija | 39 |

KAZALO

| | | |
|----------|---|-----------|
| 5 | Analiza rezultatov | 43 |
| 5.1 | Testiranje trgovalne strategije | 44 |
| 5.2 | Kakovost trgovalne strategije | 51 |
| 5.3 | Analiza gruĉ | 54 |
| 5.4 | Testiranje napovednega modela | 58 |
| 5.5 | Statistiĉna primerjava | 78 |
| 6 | Sklepne ugotovitve | 85 |
| | Literatura | 89 |
| A | Podrobni rezultati modela W2V | 93 |

Seznam uporabljenih kratic

| kratica | angleško | slovensko |
|-------------|---------------------------------------|--|
| MA | Moving Average | tekoče povprečje |
| MACD | Moving Average Convergence Divergence | tekoče poprečje konvergence divergence |
| OHLC | Open, High, Low, Close | začetna, najvišja, najnižja, zaključna cena |
| W2V | Word2Vec | Word2Vec |
| EMA | Exponential Moving Average | uteženo tekoče povprečje |
| NLP | Natural Language Processing | obdelava naravnega jezika |
| TA | Technical Analysis | tehnična analiza |

Povzetek

Naslov: Tvorba jezika japonskih svečnikov in uporaba NLP algoritma Word2Vec za napovedovanje trendov gibanja vrednosti delnic

V magistrskem delu rešujemo problem napovedovanja prihodnjih trendov vrednostnih papirjev s pomočjo strojnega učenja. Predstavimo nov inovativen model napovedovanja, ki temelji na uporabi japonskih svečnikov ter na NLP algoritmu Word2Vec. V delu pokažemo, da je možno iz zaporedja japonskih svečnikov tvoriti preprost jezik, katerega slovar je v primerjavi z naravnim jezikom sicer precej omejen. Z obdelavo Word2Vec sistem naučimo kontekst posameznih besed ter ta kontekst uporabimo pri gradnji napovedi. Napovedni model testiramo s simulatorjem trgovanja, ki v svojem delovanju upošteva tudi stroške trgovalne provizije. Rezultate predlaganega napovednega modela primerjamo s tremi osnovnimi modeli: *Kupi in zadrži*, *Tekoča povprečja* ter *MACD*. V analizi pokažemo, da predlagani napovedni model v okviru zastavljene trgovalne strategije znotraj testnega obdobja ustvari dobiček ter deluje mnogo bolje od prej omenjenih napovednih modelov. Delovanje modela preverimo tudi v validacijskem obdobju, kjer ravno tako dosežemo zadovoljive rezultate.

Ključne besede: borza, NLP, Word2Vec, japonski svečniki, strojno učenje.

Abstract

Title: The formation of Japanese candlesticks language and using NLP algorithm Word2Vec for shares trend forecasting

Our primary objective is predicting the future trends for stock market data with machine learning. We present a new innovative prediction model, based on centuries old Japanese candlesticks and modern NLP algorithm Word2Vec. Suggested model constructs a simple language of Japanese candlesticks with a very limited vocabulary. In the following steps the prediction model uses Word2Vec to discover semantic context of each word within the language vocabulary. To test the model we develop a simulation tool that takes into account the most important aspect of stock market trading – trade fees. To compare the success of the suggested prediction model we also develop simple TA models such as: Buy and Hold, Simple Moving Averages and MACD. Analysis of the results show the superiority of suggested prediction model over previously mentioned models in the test data-set. Additional testing is done with validation data-set in order to verify the results.

Keywords: stock market, NLP, Word2Vec, candlesticks, machine learning.

Poglavje 1

Uvod

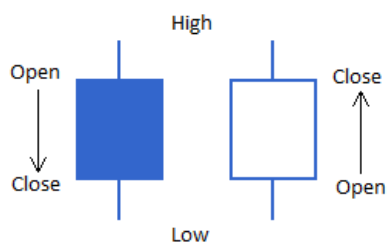
Napovedovanje prihodnjih vrednosti ali trendov gibanj vrednosti delnic je in je bilo od nekdaj zanimivo področje tako za same investitorje kot tudi raziskovalce. Čeprav so se reševanja problema napovedovanja vrednosti delnic lotili mnogi, je število uspešnih, objavljenih raziskav relativno majhno. To je seveda na nek način razumljivo, saj si prav nihče ne želi objaviti postopka, kako rešiti enega najbolj donosnih problemov.

Načinov in pristopov k reševanju problema je ogromno, v grobem pa se delijo na tehnično analizo (angl. *technical analysis*) ter temeljno analizo (angl. *fundamental analysis*). Temeljna analiza vključuje podrobno analizo poslovanja in rasti družbe, kjer se investitorji odločajo na podlagi poslovnih izidov, letnih poročil kot tudi poljubnih novic o poslovanju podjetja. Poskusi avtomatizacije trgovanja, ki temelji na podlagi obdelavi nestrukturiranih besedil [1, 2, 3], kot so novice, poročila ali objave mnenj posameznikov, temeljijo na uporabi NLP (angl. Natural Language Processing) algoritmov. Eden izmed zanimivih NLP algoritmov je *Word2Vec* [4], ki poleg svoje preprostosti prinaša tudi nadvse spodbudne rezultate v smislu razumevanja konteksta besed.

Drug način avtomatizacije trgovanja na borzi, napovedovanje prihodnjih vrednosti oziroma trendov, temelji zgolj na podlagi zgodovinskega spreminjanja cene in na tehnični analizi. Zagovorniki trdijo, da so vse potrebne

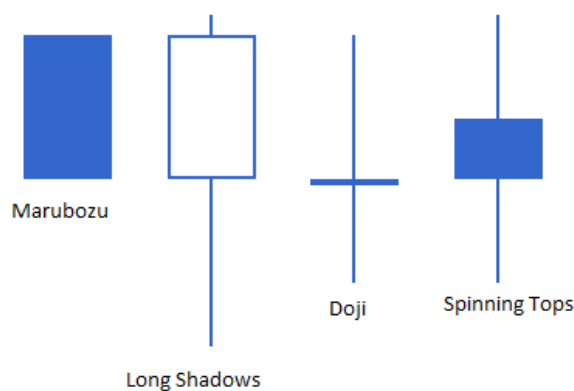
informacije vključene v samo ceno, da se zgodovina ponavlja ter da se cene gibljejo na podlagi trendov [5]. Pogosta tehnika, ki se uporablja že od leta 1700 dalje [6] in ki jo uporabljajo mnogi raziskovalci še danes, je tehnika napovedovanja s pomočjo japonskih svečnikov [5, 7, 8, 9, 10].

Japonske svečnike je razvil Munehisa Homma v 18. stoletju. Homma je obogatel s trgovanjem riža, njegova skrivnost pa je bila ravno uporaba japonskih svečnikov za analizo stanja trga in napovedovanja prihodnjih cen [6]. Japonski svečniki so primarno orodje za vizualizacijo spremembe cene v nekem časovnem obdobju. Za izris japonskega svečnika potrebujemo štiri ključne podatke: začetno ceno, najvišjo ceno, najnižjo ceno ter zaključno ceno. Ta četverček poimenujemo OHLC, ki je kratica za njihova angleška imena (angl. *Open*, *High*, *Low*, *Close*). Na Sliki 1.1 je podan primer izrisa japonskega svečnika in notacija, ki jo uporabljamo v okviru tega dela. Poznamo mnogo vrst japonskih svečnikov, na Sliki 1.2 podamo zgolj nekaj bolj pogosto predstavljenih.



Slika 1.1: Interpretacija japonskih svečnikov

Napovedno moč japonskih svečnikov v sorodni literaturi avtorji izkoriščajo predvsem na ravni ekspertnega znanja in pravil, določenih na podlagi preteklih vzorcev. Tako na primer Tsung-Hsun Lu in Yung-Ming Shiu [8] preučita 24 dvodnevni vzorcev japonskih svečnikov in njihovo napovedno moč na Tajvanski borzi. Avtorja pokažeta, da lahko določene vzorce uporabimo za



Slika 1.2: Primeri japonskih svečnikov

uspešno napoved vrednosti v naslednjem dnevu. T. Kamo in C. Dagli sta se napovedovanja lotila malo drugače. V svojem delu [9] za napovedovanje vrednosti delnic uporabljata množico specializiranih nevronske mreže – ekspertov. Vsak ekspert tako napoveduje zgolj eno izmed vrednosti OHLC. Pridobljene vrednosti avtorja nato s pomočjo mehke logike združita v končno napoved, za merjenje uspešnosti svoje metode sta uporabila mero kvadrata napake.

Drugi avtorji, kot so Milad Jasemi et. al. [7] in Karsten Martiny [11], so japonske svečnike uporabili malo drugače. M. Jasemi je s sodelavci predstavil nov pogled na tehnično analizo. Trdijo namreč, da se vnaprejšnja izbira signalov in pravil, ki prožijo nakup ali prodajo, obnese slabše, kot če se tovrstnih signalov učimo sproti. V svojem delu so za napovedovanje uporabili nevronske mreže, kjer so za vhodne podatke izbrali določene lastnosti japonskih svečnikov, kot so na primer dolžine senc. Njihova metoda doseže 74,2-odstotno uspešno napoved signalov za nakup ali prodajo, vendar pa avtorji ne predstavijo, ali metoda doseže tudi finančno uspešnost na trgu. Karsten Martiny v svojem delu [11] razišče področje nenadzorovanega učenja za potrebe napovedovanja vrednosti finančnih kazalcev. Martiny OHLC podatke

razvršča s pomočjo gručenja, nato pa z iskanjem vzorcev na več dnevnih zaporedjih napoveduje prihodnje vrednosti. Za merjenje uspešnosti metode je avtor zgolj štel število pravilno proženih signalov za nakup oziroma prodajo.

V tem delu predstavimo nov model napovedovanja prihodnjih trendov vrednosti delnic, ki združi algoritem NLP *Word2Vec* ter stoletja staro metodo japonskih svečnikov. Predlagan napovedni model temelji na združevanju tehnične analize in NLP algoritma *Word2Vec*, ki se običajno uporablja za obdelavo nestrukturiranih besedil. Osnovna predpostavka je, da je možno iz OHLC podatkov tvoriti nekakšen poenostavljen jezik, jezik OHLC oziroma jezik japonskih svečnikov, ki ga lahko nato obdelamo z algoritmom *Word2Vec*. Prepričani smo, da je možno z obdelavo jezika japonskih svečnikov z algoritmom *Word2Vec* sistem naučiti pravil in vzorcev ter to znanje uporabiti za napovedovanje prihodnjih trendov.

Kljub mnogim razvitim modelom in tehnikam napovedovanja je smiselni prikaz uspešnosti težaven. Eden izmed ciljev tega dela je tudi razviti preprost postopek testiranja napovednih modelov, katerega rezultat je finančni uspeh oziroma donos testiranega modela.

Magistrsko delo je sestavljeno iz 6 poglavij. Najprej predstavimo izbrane delnice ter njihove podatke ter te razbijemo v tri disjunktne množice, da zagotovimo objektivnost pri testiranju. Podrobnejši opis predlaganega modela je podan v poglavju 3, rezultati napovednega modela pa so podrobno analizirani in predstavljeni v poglavju 5. V 6. poglavju sledi zaključek in sklepne ugotovitve.

Poglavje 2

Podatki

Izbira podatkov, nad katerimi izvajamo testiranje delovanja napovednega modela, je poleg modela samega najverjetneje druga najbolj pomembna točka v samem postopku. Z načrtno izbiro delnic, ki so v stalni rasti, lahko dosežemo boljše rezultate, kot jih metoda sama zagotavlja. V tem poglavju podajamo podrobnejši opis izbranih delnic ter sam vpogled v splošne trende.

Pri izbiri delnic, ki jih obravnavamo v magistrskem delu, smo želeli zadostiti naslednjim pogojem:

- nabor podatkov je dovolj velik,
- cena v izbranem obdobju ni zgolj rastla,
- poznamo poslovni model podjetja.

Delnice, ki zadostijo zgornjim pogojem, smo iskali znotraj indeksa S&P500¹, ki temelji na vrednosti 500 najuspešnejših ameriških podjetij. Iz nabora 500 podjetij smo se v okviru tega dela osredotočili na delnice podjetij Apple (*AAPL*), Microsoft (*MSFT*) ter Coca-Cola (*KO*). Podjetji Apple in Microsoft spadata v IT sektor, katerega poslovni modeli so na našem področju precej jasni in razumljivi. Ker je IT sektor v zadnjih desetletjih doživel pravi razcvet, smo v nabor podatkov dodali tudi delnice podjetja Coca-Cola, katere rast je v primerjavi z delnicami IT podjetij mnogo počasnejša.

¹<http://www.investopedia.com/terms/s/sp500.asp>

Obdobje podatkov, ki smo ga izbrali, se začne s 1. majem 2000 ter konča s 1. aprilom 2016. Za vse tri delnice smo tako pridobili podatke OHLC za 4.000 trgovalnih dni, ki smo jih razdelili v tri disjunktne množice:

- učna množica,
- testna množica,
- validacijska množica.

Razdelitev podatkov med te tri množice je preprosta. Prvih 1.000 trgovalnih dni dodelimo učni množici, naslednjih 1.000 testni, zadnjih 2.000 trgovalnih dni pa obravnavamo v okviru validacijske množice. Razdelitev na tri množice je ključna za objektivno in pravilno oceno delovanja razvitega napovednega modela, hkrati pa tudi edina primerna razdelitev za analizo časovnih vrst v kontekstu gibanja vrednosti delnic [12]. Z analizo podatkov iz preteklosti lahko hitro pridemo do pristranskosti zaradi vpogleda v prihodnost [12]. V fazi razvoja in testiranja napovednega modela smo analizirali zgolj podatke v učnih in testnih množicah. Sam vpogled v smer gibanja, v obliko grafov validacijskih množic smo opravili zgolj v okviru pisanja tega dela po zaključenem testiranju in analizi rezultatov.

2.1 Vir podatkov

Podatki, na podlagi katerih se predlagani model uči ter na podlagi katerih napoveduje, so sestavljeni iz četverčka OHLC – Open, High, Low, Close ter seveda samega datuma za potrebe risanja grafov in zagotavljanja sosledja. Omenjene podatke smo pridobili iz spletnih arhivov Google Finance². Za vir Google Finance smo se odločili predvsem zato, ker v cenah že upošteva razdelitev delnic (angl. *stock split*), do katere pride, ko se vodstvo podjetja odloči povečati število delnic, ki so na voljo. Razdelitev delnic podjetju ne doda dejanske vrednosti, temveč zgolj poveča število delnic, ki so na voljo.

²<https://www.google.com/finance>

Tako se na primer ob razdelitvi delnic v razmerju 2 proti 1 število delnic poveča za 2, vrednost posamezne delnice pa se prepolovi, saj je vrednost podjetja enaka kot pred razdelitvijo.

Za vsako delnico smo pridobili nabor podatkov:

- datum,
- začetne cene (angl. *Open*),
- najvišje cene v dnevu (angl. *High*),
- najnižje cene v dnevu (angl. *Low*),
- zaključne cene v dnevu (angl. *Close*),
- število prodanih delnic.

OHLC podatki, s katerimi v nadaljevanju upravljamo, zajemajo začetno, končno, najvišjo in najnižjo ceno na stopnji posameznega dneva. Do podrobnejših vrednosti na ravni ure ali minute znotraj posameznega dneva žal ne moremo dostopati brezplačno. Ker obdelujemo delnice, ki kotirajo na ameriških borzah, so vsi zneski v nadaljevanju izraženi v dolarjih.

2.2 Coca-Cola (NYSE:KO)

Coca-Cola velja za eno izmed bolj stabilnih podjetij, saj deluje že dobrih 130 let, z njenimi delnicami pa se trguje že od leta 1920 na New Yorški borzi. Poslovni model, na podlagi katerega deluje Coca-Cola, je preprost in zelo uspešen, kar rezultira v ogromnem zaupanju s strani velikih in malih delničarjev.

Omenili smo, da smo podatke razdelili v učno, testno in validacijsko množico. Gibanje vrednosti delnice skozi posamezno obdobje prikazuje Slika 2.1.

Iz grafov lahko razberemo, da se vrednost v testnem in validacijskem obdobju giblje precej podobno, medtem ko je v učnem obdobju njeno gibanje



Slika 2.1: Gibanje vrednosti delnic Coca-Cola v učnem, testnem in validacijskem obdobju

precej bolj turbulentno, a se zaključi s približno enako vrednostjo kot na začetku obdobja.

Analiza gibanja delnice razkrije, da v podanem obdobju 4.000 trgovalnih dni delnica pridobi na vrednosti v 1.997 trgovalnih dnevih, medtem ko vrednost delnice upade v preostalih 2.003 dnevih. Povprečna rast delnice na pozitiven dan je znašala 0,91 %, povprečen padec pa je znašal 0,86 %. V podanem obdobju je največja dnevna rast dosegla kar 13,8 % povišanje vrednosti, največji dnevni padec pa je vrednost delnice podjetja zmanjšal za 10,06 %.

2.3 Microsoft (NASDAQ:MSFT)

Druga delnica, ki jo analiziramo, je delnica podjetja Microsoft, ki kotira na ameriškem NASDAQ, ter spada v sektor IT podjetij. Ta so na začetku tega tisočletja doživela velik padec vrednosti zaradi poka “dot com” mehurčka. Gibanje vrednosti delnice v vseh treh obdobjih prikazuje Slika 2.2. Razberemo lahko, da vrednost delnic tako v učnem kot testnem obdobju pada, v

zadnjem, validacijskem obdobju pa zabeleži rast. Delnice podjetja Microsoft bodo zanimive s stališča testnega obdobja, saj nam bo delovanje napovednega modela v tem obdobju dober pokazatelj, kako se le-ta obnese, ko beležimo splošen padec vrednosti.



Slika 2.2: Gibanje vrednosti delnic Microsoft v učnem, testnem in validacijskem obdobju

Analiza gibanja vrednosti delnic podjetja Microsoft je pokazala, da je število dni rasti in padcev vrednosti podobno kot pri podjetju Coca-Cola. Delnica podjetja Microsoft pridobi na vrednosti v 1.864 trgovalnih dnevih, v preostalih pa na vrednosti izgubi. Povprečna rast na pozitiven dan je znašala 1,28 %, najvišja dnevna rast pa je znašala 18,60 %. Podobno kakor pri rasti je povprečen padec vrednosti delnice znašal 1,19 %, največji dnevni padec pa 11,71 %.

2.4 Apple (NASDAQ:AAPL)

Tretja delnica, na kateri smo testirali delovanje predlaganega napovednega modela, je delnica podjetja Apple. Apple velja za eno bolj uspešnih IT podjetij, ki beleži ogromne dobičke ter rast. Pri delnicah podjetja Apple

pričakujemo v splošnem pozitiven rezultat, saj so njihove delnice v testnem in validacijskem obdobju zabeležile ogromno rast, kar prikazuje tudi Slika 2.3. Delnice podjetja Apple kotirajo, tako kot delnice podjetja Microsoft, na ameriškem NASDAQ-u.

Delnica podjetja Apple je zanimiva s stališča, da je njena vrednost v učnem obdobju precej upadla oziroma se njena vrednost v splošnem ni veliko spreminjala – gibala se je več ali manj znotraj mej \$1 ter \$2, nato pa v testni in validacijski fazi zabeležimo ogromno rast.



Slika 2.3: Gibanje vrednosti delnic Apple v učnem, testnem in validacijskem obdobju

Število dni rasti in padcev je v podobnem razmerju kakor pri delnicah Coca-Cole in Microsofta, vrednost delnice se je povečala v 1.986 dneh od 4.000, kar je skoraj polovica. Povprečna rast je znašala 1,9 %, povprečen padec pa 1,64 %. Največja zabeležena dnevna rast v analiziranem obdobju je znašala 14,13 %, največji padec pa je delnica zabeležila v učnem obdobju, ko se je njena vrednost več kot prepolovila s kar 51,83 % padcem.

2.5 Russell Top 50

Podrobnejše testiranje predlaganega napovednega modela, ki temelji na uporabi algoritma *Word2Vec*, smo opravili na naboru delnic indeksa **Russell Top 50**³. Indeks Russell Top 50, kot nakazuje samo ime, vsebuje delnice 50 različnih podjetij, seznam katerih najdemo v Tabeli 2.1. Indeks vsebuje tudi delnice podjetij Apple, Microsoft ter Coca-Cola, ki smo jih uporabili pri dosedanjih testiranjih.

Podatke OHLC za posamezno delnico smo tudi tokrat razdelili v učno, testno in validacijsko množico. Razdelitve tokrat nismo mogli opraviti na enak način kakor poprej, kjer smo podatke razdelili v množice velikosti 1.000 za učno in testno ter 2.000 primerov za validacijsko. Nekatera podjetja iz indeksa Russell Top 50 so namreč na borzi zgolj nekaj let in vsebujejo 1.000 ali celo manj OHLC podatkov. Razdelitev smo tako opravili na malenkost drugačen način. Trgovalne podatke smo za vsako delnico razdelili v tri množice na sledeči način:

- učna množica – prvih 25% OHLC podatkov,
- testna množica – naslednjih 25% OHLC podatkov,
- validacijska množica – zadnjih 50% OHLC podatkov.

Časovno obdobje večine delnic se začne s 1. januarjem 2000, zaključí pa s 16. septembrom 2016. Podjetja, katerih delnice so se na borzah pojavila šele kasneje, smo tako v analizo vključili s prvim dnevom pojavitve.

³<http://gi.guggenheiminvestments.com/products/etf/details?productid=101>

| | |
|---------------------------------|-------------------------------|
| Verizon Communications (VZ) | AT&T (T) |
| UnitedHealth Group (UNH) | Amgen (AMGN) |
| General Electric (GE) | Celgene Corporation (CELG) |
| Comcast Corporation (CMCSA) | Cocal-Cola Company (KO) |
| McDonald's Corporation (MCD) | Allergan (AGN) |
| Qualcomm (QCOM) | Schlumberger (SLB) |
| Home Depot (HD) | Bank of America (BAC) |
| Pfizer (PFE) | Wells Fargo & Company (WFC) |
| Chevron Corporation (CVX) | United Technologies (UTX) |
| Medtronic (MDT) | Honeywell International (HON) |
| Bristol Meyers Squibb (BMY) | Boeing Company (BA) |
| Intl. Bus. Machines (IBM) | Wall-Mart Stores (WMT) |
| Apple (AAPL) | Microsoft (MSFT) |
| Berkshire Hathaway (BRKB) | Mastercard (MA) |
| Walt Disney (DIS) | Visa (V) |
| 3M Company (MMM) | Philip Morris (PM) |
| Intel Corporation (INTC) | Cisco Systems (CSCO) |
| Procter & Gamble Company (PG) | Alphabet (GOOGL) |
| Union Pacific Corporation (UNP) | Johnson & Johnson (JNJ) |
| Merck & Company (MRK) | Exxon Mobil (XOM) |
| Altria Group (MO) | Amazon.com (AMZN) |
| Abbvie (ABBV) | Gilead Sciences (GILD) |
| Oracle (ORCL) | Facebook (FB) |
| Citigroup (C) | CVS Health (CVS) |
| Pepsico (PEP) | JP Morgan Chase & Co (JPM) |

Tabela 2.1: Seznam delnic indeksa Russel Top 50

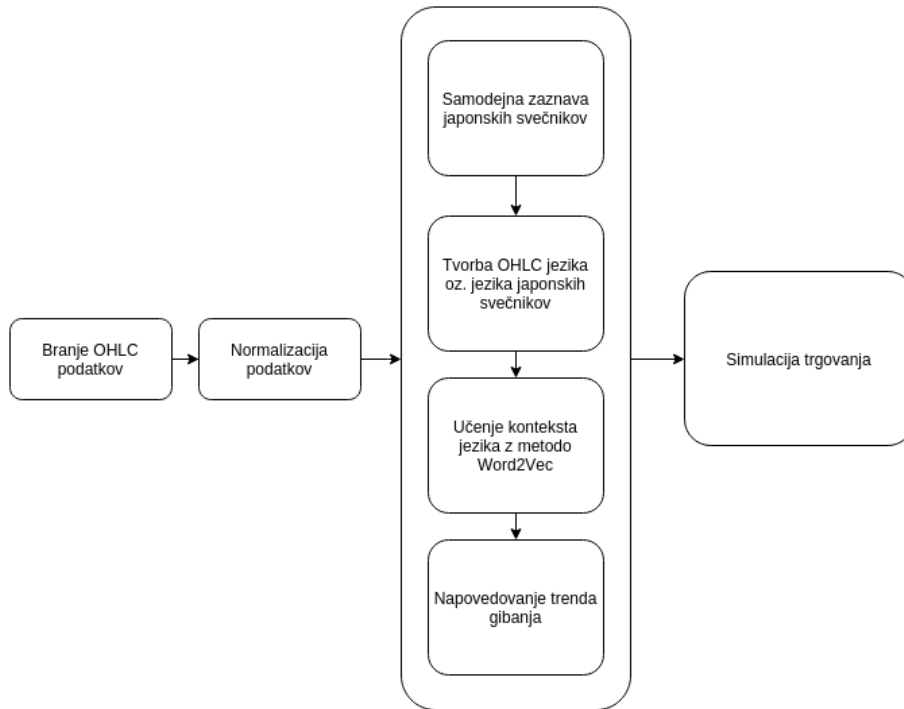
Poglavje 3

Opis napovednega modela

V nadaljevanju podajamo podrobnejši opis in predlog novega napovednega modela, ki združuje množico različnih metod strojnega učenja na nov in inovativen način. Osnovna predpostavka, na kateri je zgrajen predlagani napovedni model, je, da japonski svečniki niso le dobro orodje za grafični prikaz gibanja vrednosti *OHLC* podatkov, temveč vsebujejo tudi napovedno moč [9, 7, 8, 10]. Osnovna ideja se je porodila ob branju zgodbe o Munehisu Hommi, ki je s pomočjo interpretacije japonskih svečnikov obogatel že davnega leta 1700 [6]. Predlagani model predstavi popolnoma nov pristop k obravnavi japonskih svečnikov. Predpostavka, na kateri gradimo, je, da različni vzorci japonskih svečnikov lahko napovedo prihodnje gibanje vrednosti. Ključni so torej različni vzorci in zaporedja, v katerih se japonski svečniki pojavljajo. Japonski svečniki so v tem pogledu kakor jezik – tudi ta je sestavljen iz besed, zaporedja oziroma vzorci le-teh pa tvorijo stavke in povedi, ki vsebujejo neki globlji pomen. Predlagan napovedni model se močno naslanja na zgornjo primerjavo japonskih svečnikov z naravnim jezikom. Napovedi gradimo s pomočjo transformacije *OHLC* podatkov v nekakšen poenostavljen jezik japonskih svečnikov. Pri gradnji napovedi pridobljeni jezik obdelamo z NLP algoritmom *Word2Vec* [4], kjer želimo računalnik naučiti določenih lastnosti in zakonitosti predlaganega jezika z namenom uporabiti odkrite in naučene lastnosti ter zakonitosti pri samem napovedovanju priho-

dnjih trendov smeri gibanja vrednosti delnic.

V nadaljevanju podajamo podrobnejši opis vsakega izmed korakov, ki so potrebni za delovanja predlaganega napovednega modela. Diagram delovanja predlaganega napovednega modela je podan na Sliki 3.1.



Slika 3.1: Diagram delovanja napovednega modela

3.1 Samodejna zaznava japonskih svečnikov

Kot omenjeno na začetku tega poglavja, predlagani napovedni model temelji na japonskih svečnikih. Slabost mnogih modelov, ki temeljijo na japonskih svečnikih, je njihova odvisnost od vnaprejšnje definicije oblik japonskih svečnikov. Bodisi so to točne definicije japonskih svečnikov bodisi njihovih lastnosti. Zahteva po vnaprejšnji definiciji s strani ekspertov ni zaželena lastnost, saj je delovanje modela tako omejeno s količino in kakovostjo vnaprej definiranih lastnosti japonskih svečnikov [11].

Ključno za delovanje predlaganega napovednega modela bo torej samodejna zaznava japonskih svečnikov. Ker ne želimo, da bi bil naš model omejen s pravili, določenimi s strani ekspertov, moramo v tem koraku uporabiti metodo nenadzorovanega strojnega učenja za gručenje – *KMeans*. Poleg dejstva, da bi bila vnaprejšnja definicija skupin japonskih svečnikov precej zamudna, bi med učnimi ali testnimi podatki lahko našli primere, ki ne bi ustrezali naši definiciji, kar bi povzročilo težave pri napovedovanju. Dodatna motivacija za samodejno razvrščanje japonskih svečnikov je tudi uspeh avtorjev v [11] in [7]. Samodejno razvrščanje je potrebno tudi zaradi želje po razvoju jezika japonskih svečnikov. Razviti jezik japonskih svečnikov bo že tako vseboval precej omejen nabor besed, z ročno definicijo pa bi ta nabor še dodatno zmanjšali. Dodatno smo z uporabo nenadzorovanega učenja pridobili še možnost nastavljanja parametra števila besed oz. japonskih svečnikov.

Sam postopek gručenja je prikazan na izseku kode 3.1. Pri razvoju smo uporabili programski paket **sklearn**¹, ki vsebuje množico razvitih algoritmov strojnega učenja. Pomembno je, da v sklopa razvoja upoštevamo, da algoritem *KMeans* centroide v fazi učenja določa naključno, zato lahko z enakimi parametri ob ponovnem zagonu pridobimo različne razrede. Ker želimo zagotoviti ponovljivost vseh rezultatov, nastavimo algoritmu parameter **random_state**.

¹<http://scikit-learn.org/stable/>

```
1 from sklearn.cluster import KMeans
2
3 # Ucenje KMeans z ucno mnozico OHLC podatkov
4 kmeans = KMeans(n_clusters=50, random_state=42)
5 kmeans.fit(train_ohlc_data)
```

Koda 3.1: Postopek gručenja

3.1.1 KMeans

Algoritem *KMeans* spada v kategorijo nenadzorovanih metod strojnega učenja, saj za svoje delovanje ne potrebuje vnaprej podanih pravilnih trditev, na podlagi katerih bi izluščil pravila za nadaljnje napovedovanje. Algoritem na podlagi želenega števila gruč, parametra k in števila podatkov n v večdimenzionalnem prostoru poišče k centroidov tako, da minimizira skupno kvadratno razdaljo med vsako točko in njenim najbližjim centroidom [13]. Funkcija, ki jo algoritem minimizira, je podana v Formuli 3.1, kjer je x podatek iz množice X .

$$\Phi = \sum_{x \in X} \min_{c \in C} \|x - c\|^2 \quad (3.1)$$

Samo delovanje algoritma je preprosto, saj ima zgolj tri korake. V prvem koraku algoritem naključno izbere začetne centroide z izborom k primerov iz množice podatkov X . V drugem koraku algoritem vsak primer na podlagi razdalje do centroida pripiše eni izmed gruč. V tretjem koraku se novi centriodi izračunajo s povprečenjem vrednosti primerov znotraj trenutnih gruč. Algoritem poljubnokrat ponovi koraka dva in tri, dokler ne dobimo oblikovanih gruč, kar pomeni, da se centriodi več ne spreminjajo [13]. Potencialna slabost algoritma *KMeans* je njegova nagnjenost k iskanju konveksnih gruč. Tako na primer slabo zazna gruče, katerih oblika je podolgovata ali drugih nepravilnih oblik.

Silhueta

Kakovosti rezultatov metode *KMeans* pravzaprav ni tako preprosto določiti. Pravilnih vrednosti za posamezen *OHLC* podatek nimamo, tako da se moramo za oceno kakovosti nasloniti na metodo samo. Mera, ki smo jo uporabili v tem delu in katere opis podajamo v nadaljevanju, se imenuje Silhueta.

Izračun Silhuete je zelo preprost. Za vsak primer znotraj gruče izračunamo njegovo povprečno razdaljo **a** do vseh ostalih primerov znotraj iste gruče. Nato izračunamo še povprečno razdaljo **b** od primera do vseh ostalih primerov v naslednji najbližji gruči.

Silhueta za posamezen primer se nato izračuna po formuli 3.2. Najboljša vrednost ocene Silhueta je vrednost 1, najslabša pa ocena -1. Skupna ocena se izračuna s povprečenjem silhuet posameznih primerov. Podrobnejši opis Silhuete se nahaja v [14].

$$s = \frac{b - a}{\max(a, b)} \quad (3.2)$$

3.1.2 Normalizacija OHLC podatkov

Naši japonski svečniki so predstavljeni v obliki *OHLC* vrednosti. Na Sliki 3.2 si lahko ogledamo prikaz surovih podatkov, ki v takšni obliki še niso primerni za nadaljnjo obdelavo. Japonski svečniki so na grafu vertikalno pozicionirani glede na njihovo relativno vrednost.

Podatke moramo pred nadaljnjim delom najprej normalizirati. Normalizacijo opravimo z deljenjem *OHLC* - *Open*, *High*, *Low*, *Close* podatkov z vrednostjo *Open* [11], kot prikazano na enačbi 3.3 .

$$\text{norm}(\text{OHLC}) = \frac{H}{O}, \frac{L}{O}, \frac{C}{O} \quad (3.3)$$

Poglejmo si prikaz podatkov s Slike 3.2 po postopku normalizacije na sliki 3.3. *OHLC* podatki zdaj ne vsebujejo več podatkov o absolutnih vrednostih na določen dan, temveč podajajo zgolj obliko japonskega svečnika.



Slika 3.2: Surovi OHLC podatki

Pri samem postopku normalizacije smo opazili, da je naš vir podatkov, Google Finance, nepopoln. Med pregledovanjem podatkov smo našli nekaj 10 takšnih trgovalnih dni, kjer je manjkala ena izmed potrebnih vrednosti iz četverčka *OHLC*. V nadaljevanju smo nepopolne trgovalne dneve izključili iz obdelave.

3.1.3 Število japonskih svečnikov

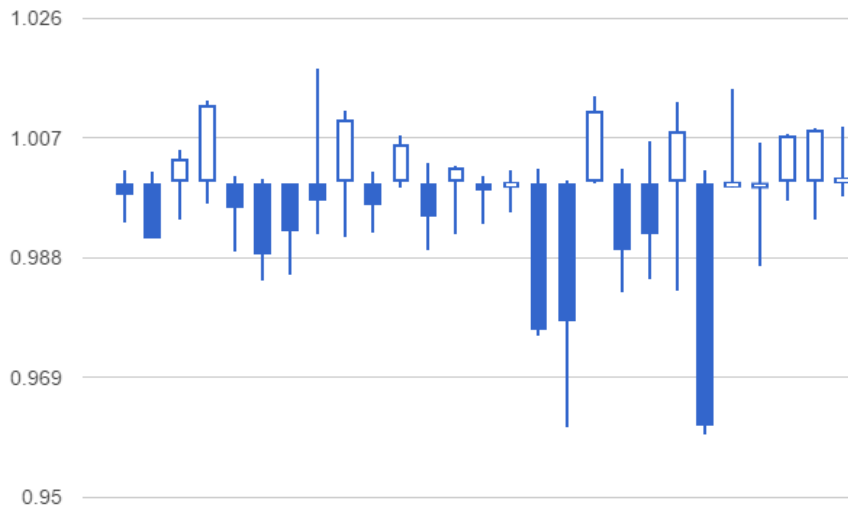
Algoritem *KMeans* sprejme zgolj nekaj parametrov – kot smo videli na izseku kode 3.1, je obvezen zgolj parameter **n_clusters**, s katerim določimo, v koliko gruč želimo razporediti podatke.

Ker pravilna razporeditev *OHLC* podatkov v skupine japonskih svečnikov oziroma gruče ni znana, je treba kakovost ocene gruč podati s pomočjo samega algoritma gručenja. Poglejmo si rezultate *Silhouette* za različne vrednosti parametra **n_clusters** v Tabeli 3.1.

Ob analizi rezultatov v Tabeli 3.1 vidimo, da bi bila najboljša razdelitev v dve gruči. Ker bodo gruče predstavljale osnovo za naš jezik, nam takšna

| Št. gruč | Silhueta |
|----------|----------|
| 2 | 0,43 |
| 3 | 0,41 |
| 4 | 0,38 |
| 5 | 0,37 |
| 6 | 0,34 |
| 7 | 0,33 |
| 8 | 0,34 |
| 9 | 0,32 |
| 10 | 0,28 |
| 15 | 0,29 |
| 20 | 0,30 |
| 25 | 0,29 |
| 30 | 0,30 |
| 35 | 0,28 |
| 40 | 0,30 |
| 50 | 0,29 |
| 60 | 0,28 |
| 70 | 0,29 |
| 80 | 0,29 |
| 90 | 0,30 |
| 100 | 0,30 |

Tabela 3.1: Silhueta za različne parametre **n_clusters**



Slika 3.3: OHLC podatki po normalizaciji

razdelitev seveda nikakor ne ustreza. V okviru OHLC jezika si želimo 10 in več različnih besed, tudi ročno definiranih japonskih svečnikov je mnogo več kot zgolj dve skupini [6], že pri hitri analizi japonskih svečnikov na Sliki 3.3 vidimo več kot zgolj dve skupini.

S podatki iz Tabele 3.1 smo dobili potrditev, da z večanjem števila japonskih svečnikov oziroma števila besed v našem jeziku ne izgubimo veliko v okviru silhuetne ocene kakovosti gruč. Seveda nižja silhuetna ocena pomeni, da imamo veliko primerov, ki bi lahko pripadali tudi drugi gruči, a nas to ne moti pretirano. V naravnem jeziku imamo prav tako kopico besed, ki so si med seboj zelo podobne, a nosijo popolnoma drugačen pomen, npr. “oddati” in “oddaja”.

Podrobnejša analiza gručenja in primernosti algoritma *KMeans* je podana v poglavju 5.3.

3.2 Tvorba OHLC jezika

Podatke OHLC smo uspešno normalizirali in razvili sistem samodejne zaznave japonskih svečnikov. Naslednji korak je tvorba OHLC jezika. Osnovno enoto našega jezika predstavljajo posamezne vrednosti *Open*, *High*, *Low* in *Close*. Ker je nabor le-teh praktično neskončen, je neskončen tudi nabor besed. Nabor možnih besed smo z gručenjem drastično zmanjšali na zgolj nekaj deset različnih besed in tako zelo poenostavili strukturo OHLC jezika. Treba je poudariti, da je predlagana struktura jezika zgolj ena izmed možnosti ter da z definicijo jezika vplivamo tudi na samo delovanje napovednega modela.

Velikost slovarja predlaganega OHLC jezika bo nastavljiva s pomočjo parametra **n_clusters**. Iz pridobljenih besed moramo sestaviti še stavke. Besede v stavkih, tako kot pri naravnem jeziku, pridobijo nek kontekst, s pomočjo katerega interpretiramo njihov pomen.

3.2.1 Stavki

Iz besed, ki so rezultat obdelave z algoritmom *KMeans*, v drugem koraku tvorimo stavke. Ker zaporedja japonskih svečnikov vsebujejo določeno napovedno moč [6, 8], smo le-ta vzeli za osnovo pri tvorbi stavkov. Kot že omenjeno, predlagani model ne vsebuje nikakršnih vnaprej definiranih pravil, zato s tvorbo stavkov zgolj ustvarimo zaporedja vzorcev, ki jih uporabimo pri nadaljnji obdelavi.

Na izseku 3.2 podajamo kodo, ki skrbi za tvorbo stavkov. Rezultat gručenja iz prejšnjega koraka je seznam oznak za posamezno gručo. Oznake gruč oziroma japonskih svečnikov so zaporedne številke gruč. Pri tvorbi stavkov določimo zgolj želeno dolžino stavkov, ki jih ustvarimo s pomočjo drsečega okna. Iz seznama 1.000 OHLC podatkov in dolžino stavkov 10 v učnem obdobju tako pridelamo 991 stavkov.

```
1 def construct_sentences(clusters, sentence_length):
2     """
3     Construct sentences from clusters - use sliding
4     window.
5     :param clusters:
6     :param sentence_length
7     :return:
8     """
9     return [map(str, cls) for cls in
10             window(clusters, sentence_length)]
11
12 def window(seq, n=2):
13     """
14     Returns a sliding window (of width n) over data from the
15     iterable
16
17     :param seq:
18     :param n:
19     :return:
20     """
21     it = iter(seq)
22     result = tuple(islice(it, n))
23     if len(result) == n:
24         yield result
25     for elem in it:
26         result = result[1:] + (elem,)
27         yield result
```

Koda 3.2: Tvorba OHLC stavkov

Čeprav je OHLC jezik na prvi pogled videti zelo preprost, je nabor možnih vzorcev oziroma stavkov огromen. Vzemimo za primer definicijo jezika, kjer ima slovar 20 besed ter gradimo stavke dolžine 10. Število permutacij, ki jih lahko ustvarimo s takim jezikom, znaša kar 670.442.572.800 različnih stavkov.

Pri tvorbi stavkov smo dodali še en parameter, ki lahko vpliva na kakovost napovednega modela. Ta parameter smo poimenovali **sentence_length**,

določa pa dolžino stavkov.

S tem smo zaključili definicijo OHLC jezika. Število besed, ki jih bo jezik vseboval, smo definirali v prvem koraku, v drugem pa pridobljene besede zgolj združimo v zaporedja stavkov oziroma vzorcev, ki bodo služili kot učni podatki v naslednjih fazah.

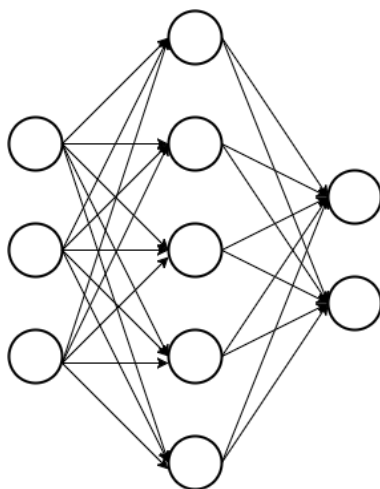
3.3 Obdelava z NLP algoritmom Word2Vec

V tretjem koraku predlagani napovedni model opravi analizo stavkov in učenje konteksta posameznih besed OHLC jezika – japonskih svečnikov. Iz vzorcev, ki so nastali na podlagi združevanja zaporednih OHLC podatkov v stavke, model v tem koraku zgradi kontekst jezika, ki ga nato v naslednjem koraku uporabi za grajenje napovedi. Sistem se tako sam, na podlagi množice zgodovinskih podatkov oziroma obstoječih vzorcev, nauči konteksta ter ga sproti z novim znanjem tudi osvežuje. Obdelavo stavkov opravimo s pomočjo NLP algoritma *Word2Vec* [4].

3.3.1 Word2Vec

Algoritem *Word2Vec*, ki ga podrobneje opisujejo Mikolov in sodelavci [4], iz ogromnega nabora besedil pridobi vektorje posameznih besed, ki vsebujejo mnogo lingvističnih pravil in vzorcev. Algoritem je tako zmožen linearne translacije, ki po translaciji vektor(“Madrid”) - vektor(“Spain”) + vektor(“France”) rezultira v vektorju, ki je bližje vektorju besede “Paris” kot katerikoli drugi besedi [4].

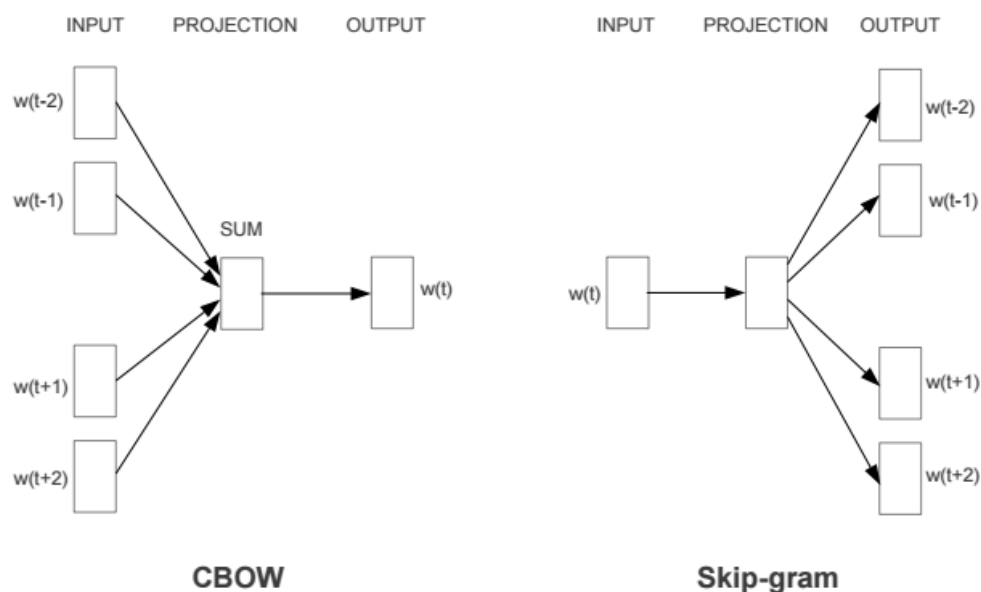
Algoritem temelji na uporabi nevronske mreže v fazi učenja. Nevronska mreža je metoda strojnega učenja, ki simulira delovanje človeških oziroma živalskih možganov. Podobno kakor v možganih tudi nevronska mreža sprejme nabor vhodnih podatkov, ki nato potujejo skozi arhitekturo nevronske mreže. Arhitektura nevronske mreže je sestavljena iz vhodnega nivoja, vmesnih skritih nivojev ter izhodnega nivoja. Nevroni so med nivoji povezani s povezavami, kot prikazuje Slika 3.4.



Slika 3.4: Primer arhitekture Nevronske mreže

Nevronska mreža, ki jo uporablja algoritem *Word2Vec*, je relativno preprosta, saj vsebuje zgolj en skriti nivo. V fazi učenja nevronska mreža na vhod prejme trenutno besedo ter na podlagi uteži Θ izračuna vektorje sosednjih besed na izhodnem nivoju. Nato s pomočjo pravih vrednosti vektorjev sosednjih besed in funkcije napake izračuna napako napovedi ter popravi uteži Θ . Tak postopek učenja imenujemo vzvratno razširjanje napake (angl. *backpropagation*). Mikolov in sodelavci so takšno arhitekturo poimenovali *Skip-gram* [4, 15]. Skip-gram avtorji predstavijo kot alternativo postopku vreča besed – CBOW (*Continuous Bag of Words*), ki iz množice sosednjih besed izračuna vektor besede. Skip-gram, kot smo opisali zgoraj, deluje ravno obratno – iz podane besede dobimo nabor sosednjih besed, ki se pojavljajo v njeni okolici. Razliko med CBOW in Skip-gram najlepše prikaže Slika 3.5.

Algoritem *Word2Vec* se izkaže kot odlično orodje za analizo naravnega jezika, kar smo videli že v prvem primeru, kjer z vektorskim seštevanjem in odštevanjem besed Madrid - Spain + France pridemo do vektorja besede Paris. Avtorji so v fazi učenja algoritma uporabili nabor časopisnih novic, ki je vseboval 1 milijardo besed, celoten besedni zaklad pa je znašal 692 tisoč različnih besed. Za dobro delovanje algoritma je ključna dovolj velika



Slika 3.5: Primerjava arhitektur CBOW in Skip-gram [15]

učna množica. Z večanjem učne množice so avtorji pokazali, da je algoritem *Word2Vec* sposoben zaznati tudi fraze, kot so “German” + “airlines” = “Lufthansa”, česar z manjšo učno množico niso dosegli [4].

3.3.2 Implementacija

Japonske svečnike, ki predstavljajo besede v OHLC jeziku, smo združili v stavke, kjer ima vsak japonski svečnik nabor sosednjih svečnikov. Z učenjem nevronske mreže z algoritmom *Word2Vec* tako pridobimo model, ki vsak japonski svečnik pretvori v njegovo vektorsko predstavitev.

Pri razvoju smo uporabili knjižnico Gensim², ki je napisana za programski jezik Python in že vsebuje optimizirano implementacijo *Word2Vec*. Sama implementacija učenja *Word2Vec* je podana v izseku kode 3.3.

²<https://radimrehurek.com/gensim/>

```
1 w2v_model = Word2Vec(sentences=sentences,
2                       size=100,
3                       window=3,
4                       min_count=5,
5                       workers=1,
6                       seed=42)
```

Koda 3.3: Učenje *Word2Vec*

Algoritem sprejme sledeč nabor parametrov:

1. **sentences** – seznam stavkov iz drugega koraka,
2. **size** – dimenzija vektorjev, v katere preoblikujemo besede,
3. **window** – število sosednjih besed, ki vplivajo na kontekst besede,
4. **min_count** – minimalno število pojavitev besede v učni množici,
5. **workers** – število uporabljenih procesov za pospešitev učenja,
6. **seed** – seme naključnega generatorja števil. Nastavimo, da zagotovimo ponovljivost rezultatov.

Parametra **size** in **window** veljata za zelo pomembna, saj z njima določimo dimenzijo prostora, v katerega preslikujemo, ter število sosednjih besed, ki vpliva na kontekst trenutne besede. V naši implementaciji nastavimo vrednost parametra **window** na enako vrednost kakor dolžino stavkov – parameter **sentence_length**. Vrednost parametra **size** smo nastavili na privzeto vrednost 100.

3.4 Napovedovanje

Dosedanji koraki so uspešno preslikali OHLC podatke v eno izmed gruč japonskih svečnikov, združili le-te v zaporedja ter predlagani model naučili konteksta posameznih japonskih svečnikov. Predlagani model bi v tej točki moral biti sposoben s pomočjo konteksta, ki se ga je naučil, in s pomočjo zgodovinskih podatkov že tvoriti napovedi za prihodnost. Predlagan način napovedovanja je zgolj eden izmed možnih načinov, kako lahko pridobljeno znanje izkoristimo za pridobivanje napovedi.

Naša želja je, da napovedni model na podlagi vhodnih podatkov na izhodu proži eno izmed treh akcij:

- kupi,
- prodaj,
- ne naredi ničesar.

Napovedi smo gradili s pomočjo nadzorovanega strojnega učenja, kjer smo algoritmu na vhodu posredovali učne primere in nabor pravilnih rezultatov, ki smo jih pridobili z analizo podatkov. Opisani problem spada v kategorijo klasifikacijskih algoritmov strojnega učenja, kjer imamo na vhodu primer, ki ga moramo razvrstiti v eno izmed treh zgoraj opisanih skupin (kupi, prodaj, ne naredi ničesar). Odločili smo se za uporabo *SoftMax* algoritma.

3.4.1 SoftMax

V nasprotju s tem, kar bi morda lahko sklepali iz njenega imena, se SoftMax regresija uporablja za večrazredno klasifikacijo primerov.

Algoritem SoftMax za podan primer x izračuna verjetnost $p(y=j | x)$, za vsak $j = 1, \dots, k$, pripadnosti primera za vsakega izmed k možnih razredov. Hipoteza algoritma SoftMax je predstavljena v Formuli 3.4, kjer so θ parametri modela. Funkciji, ki ju še potrebujemo za implementacijo algoritma, sta funkciji izgube – *cost* ter njen odvod – podani sta v 3.5 ter 3.6.

$$h_{\theta}(x^i) = \frac{1}{\sum_{j=1}^k e^{\Theta_j^T x^i}} \begin{bmatrix} e^{\Theta_1^T x^i} \\ e^{\Theta_2^T x^i} \\ \vdots \\ e^{\Theta_k^T x^i} \end{bmatrix} \quad (3.4)$$

$$J(\Theta) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^k 1\{y^i = j\} \log p(y^i = j \mid x^i; \Theta) \quad (3.5)$$

$$\nabla J(\Theta) = -\frac{1}{m} \sum_{i=1}^m [x^i (1\{y^i = j\} - p(y^i = j \mid x^i; \Theta))] \quad (3.6)$$

Implementacije formul 3.4, 3.5 in 3.6 so podane v izseku kode 3.4.

```

1 def h(self, X, theta_matrix):
2     theta_x = np.dot(X, theta_matrix.T)
3     hypothesis = np.exp(theta_x)
4     probabilities = (hypothesis.T
5                     / np.sum(hypothesis, axis=1)).T
6
7     return probabilities
8
9 def cost(self, theta, X, Y, lambda_):
10    theta_matrix = theta.reshape(Y.shape[1], X.shape[1])
11
12    probabilities = self.h(X, theta_matrix)
13
14    cost = -(np.sum(np.multiply
15                  (Y, np.log(probabilities))) / X.shape[0]) +\
16            (0.5 * lambda_ * np.sum(np.multiply(theta_matrix,
17                                                  theta_matrix))
18    )
19    return cost
20

```

```
21 def grad(self, theta, X, Y, lambda_):
22     theta_matrix = theta.reshape(Y.shape[1], X.shape[1])
23     probabilities = self.h(X, theta_matrix)
24
25     grad = np.dot(X.T, (probabilities - Y)).T /
26             X.shape[0] + lambda_ * theta_matrix
27
28     return grad.ravel()
```

Koda 3.4: Implementacija algoritma SoftMax

Postopek minimizacije napake funkcije *cost* izvajamo iterativno z gradientnim spustom. V naši implementaciji algoritma SoftMax smo dodali tudi regularizacijo, ki algoritmu preprečuje, da bi se v fazi učenja preveč prilagodil podatkom.

Regularizacija

Pogosto moramo biti pri strojnem učenju previdni, da se algoritem v fazi učenja ne prilagodi podatkom v popolnosti (angl. *overfitting*).

Čezmerno prilagajanje podatkom lahko rešimo z regularizacijo. Ker se določene vrednosti θ pri čezmernem prilagajanju podatkov enormno povečajo, je naloga regularizacije kaznovati modele, kjer so te vrednosti visoke. Poznamo dva tipa regularizacije:

- **L2** – vsota kvadratov vseh θ .
- **L1** – absolutna vsota vseh θ .

V naši implementaciji smo se odločili za uporabo L2 regularizacije, ki je tudi ena izmed bolj pogosto uporabljenih regularizacij. Stopnjo regularizacije nastavljamo s parametrom λ , ki ga dodamo množici parametrov, za katere želimo v fazi učenja poiskati optimalno vrednost.

3.5 Trgovalna strategija

Algoritem SoftMax za svoje delovanje potrebuje nabor učnih podatkov, ki so razdeljeni na primere in njihove pravilne vrednosti razreda y . Omenili smo že, da želimo na podlagi vhodnih podatkov na izhodu prožiti eno izmed akcij: *kupi*, *prodaj* ali *ne naredi ničesar*.

Algoritem SoftMax bo tako torej napovedoval enega izmed treh razredov, ki jih bomo preprosto označili kot:

- Razred **0** – kupi,
- Razred **1** – prodaj,
- Razred **2** – ne naredi ničesar.

Naša naloga je posamezne OHLC podatke v učni množici opremiti s praviimi vrednostmi **0**, **1** ali **2** na podlagi vpogleda v prihodnost. Seveda moramo biti v fazi učenja pozorni, da operiramo zgolj z informacijami, ki so nam na voljo v učni množici, ter da ne posegamo po kakršnemkoli znanju ali informacijah, ki postane na voljo v času testne ali validacijske množice.

Predlagani napovedni model bo zgodovinske podatke analiziral ter pripisal primerno vrednost razreda OHLC podatkom na podlagi trgovalne strategije, ki smo jo uporabili tudi pri simulaciji trgovanja v fazi testiranja modela. Strategij nakupov oziroma prodaj delnic je praktično neskončno, saj lahko pravila poljubno definiramo. Izbira oziroma določitev trgovalne strategije je zelo pomembna. Ena izmed naših hipotez je, da s slabo trgovalno strategijo kljub odličnemu napovednemu modelu ne moremo pričakovati dobrih donosov.

3.5.1 Trgovalna provizija

Trgovanje z delnicami na borzi seveda ni brezplačno. Pri samem nakupu smo poleg tržne vrednosti delnic borznoposredniški hiši dolžni plačati tudi znesek v višini trgovalne provizije. Načini obračunavanja ter višine provizij

se med različnimi borznimi posredniki drastično razlikujejo. Poznamo modele, kjer plačamo provizijo v fiksnem znesku, v določenem deležu celotnega zneska ali s kombinacijo obeh. Za fizične osebe je pogost model fiksnega obračunavanja provizije v zneskih med 10 in 20 dolarji na transakcijo. To pomeni, da provizijo plačamo 2-krat, prvič ob nakupu delnic, drugič pa ob njihovi prodaji.

V primeru \$1.000 začetnega vložka pri proviziji \$20 pridemo do skupnega zneska provizije \$40, ki ga plačamo ob nakupu ter kasneje ob prodaji delnic. Preračunano v odstotke to pomeni, da je vrednost delnice v času našega lastništva morala zrasti za vsaj 4 %, če želimo pokriti samo stroške trgovanja.

Štiriodstotna rast delnic je zgodovinsko gledano kar precejšnja. Zgodovinski podatki gibanja vrednosti indeksa S&P500, ki vključuje delnice kar 500 ameriških podjetij, kažejo, da je povprečen donos med letoma 1950 in 2009, po prilagoditvi za inflacijo in dividendami, znašal zgolj 7 %³.

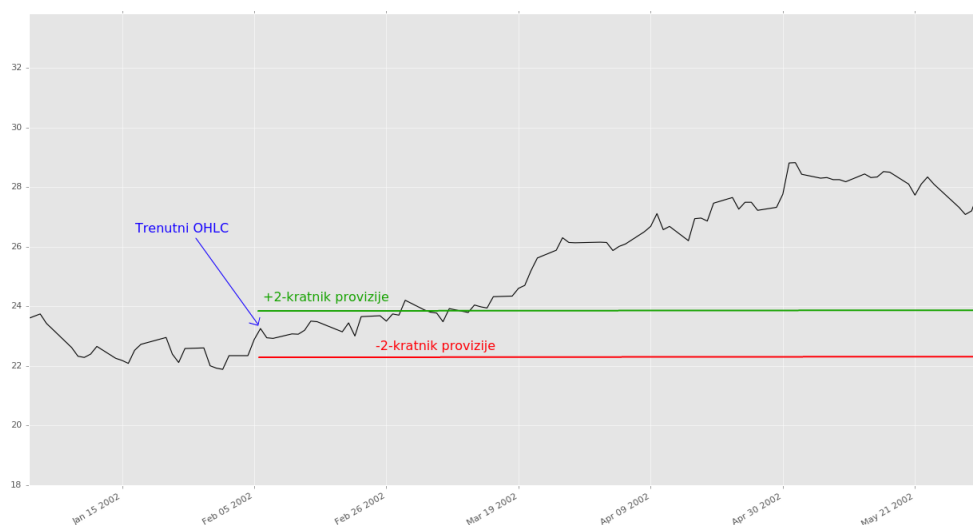
Seveda ne gre zanemariti tudi prednosti fiksne provizije, saj le-ta postane zanemarljivo majhna pri višjih vložkih. Znesek trgovalne provizije, ki smo ga določili v okviru tega dela, smo nastavili na \$15.

3.5.2 Označevanje OHLC podatkov z ustreznim razredom

Pri trgovanju želimo z vsako akcijo maksimizirati dobiček, v primeru izgub pa seveda minimizirati le-te. Poleg samega gibanja vrednosti delnice moramo pri vsaki akciji upoštevati tudi vpliv provizije. Poenostavljeno to pomeni, da moramo ob napovedi rasti pred samim nakupom upoštevati tudi znesek provizije ter ponovno oceniti donosnost nakupa v dani točki. Podobno velja tudi v primeru padca vrednosti: pred odločitvijo o prodaji moramo upoštevati, da poleg same izgube zaradi padca vrednosti plačamo dodatno še znesek v višini provizije.

³<http://www.thesimpledollar.com/where-does-7-come-from-when-it-comes-to-long-term-stock-returns/>

Hitro postane jasno, da provizija igra pomembno vlogo pri samem delovanju trgovalne strategije ter posledično tudi napovednega modela. Predlagani napovedni model za podani dan poda priporočilo akcije: kupi, prodaj, ne naredi ničesar. V fazi učenja se model v učni množici za vsak posamezen dan sprehodi določeno število dni v prihodnost ter analizira samo gibanje vrednosti delnice. V primeru, da je vrednost delnice v prihajajočih dneh presegla 2-kratnik provizije, trenutni dan označimo z razredom **0** – kupi delnice. Logika za našim razmišljanjem je sledeča – nakup se nam splača, saj bomo kljub nakupni proviziji in prodajni proviziji čez n dni še vedno ustvarili dobiček. Na Sliki 3.6 podajamo grafično ponazoritev označevanja OHLC podatka za primer razreda **0**. V določenem dnevu izračunamo količino delnic, ki jo lahko kupimo z višino osnovnega kapitala, ter pogledamo, kakšno vrednost mora delnica v prihajajočih dneh doseči, da se nam nakup splača (*zelena črta*).



Slika 3.6: Primer označevanja OHLC podatkov

Do sedaj smo definirali pravilo, kdaj posameznemu OHLC podatku pripišemo vrednost razreda **0**. V želji po minimiziranju izgube v primeru padca vrednosti delnice želimo delnice prodati zgolj v primeru, da zaznamo padec, ki presega 2-kratnik trgovalne provizije. Tako smo na Sliki 3.6 pridobili dve meji, na podlagi katerih pripišemo vrednost razreda **0** ali **1**. V primeru, ko

vrednost delnice v oknu, ki ga gledamo, ne preseže 2-kratnika niti ne pade pod 2-kratnik provizije, trenutnemu dnevu pripišemo vrednost razreda **2**.

Z zgornjim opisom smo predlaganemu napovednemu modelu dodali nov parameter – dolžina okna, s katerim pripišemo ustrezne vrednosti razredov. Parameter smo poimenovali **period_frame_days**. Njegov vpliv na delovanje napovednega modela podrobneje analiziramo v fazi testiranja modela.

Kdaj delnice kupimo ali prodamo, je razvidno iz samih akcij. Naša strategija ob napovedani akciji *kupi* kupi toliko delnic, kot nam dopušča stanje kapitala. Podobno ob proženem signalu *prodaj* prodamo čisto vse delnice, ki si jih trenutno lastimo. V primeru, ko napovedni model trenutni dan razvrsti v razred **2**, naredimo točno to, kar nam pove ime razreda – *ne naredi ničesar*. Delnic tako v tem primeru, če si jih lastimo, ne prodamo, v primeru, ko si jih ne lastimo pa ne kupimo. Podrobnejši potek odločanja o nakupu ali prodaji opišemo v poglavju 4.2.1.

Poglavje 4

Metrike uspešnosti

Preden se podamo v podrobnejšo analizo predlaganega napovednega modela in trgovalne strategije, moramo najprej definirati metrike uspešnosti. Definicija metrik uspešnosti se morda ne zdi tako pomembna točka v celotnem procesu, a izbira metrik definira tudi samo interpretacijo uspešnosti. Zato je poglobljen razmislek okoli izbire metrik še kako potreben in smotr.

Pogosto uporabljena metrika uspešnosti je odstotek pravilnih napovedi smeri gibanja cene [16, 7, 8]. Kljub temu, da je to pomemben podatek, lahko hitro vidimo, da s to metriko težko ocenimo delovanje kateregakoli modela ob dejanskem trgovanju. Druga metrika uspešnosti napovednega modela, ki bi jo lahko uporabili v primeru napovedovanja dejanske vrednosti delnic v prihodnosti [9], bi bila povprečje kvadratne napake – MSE (angl. *Mean Squared Error*). Vendar pa izhodni podatki predlaganega napovednega modela ne vsebujejo dejanske napovedi cene delnic v prihodnosti, temveč zgolj splošen trend (pozitiven, negativen ali stagnacija), tako da metrike MSE žal ne bomo mogli uporabiti.

Pri izboru katerekoli metrike moramo razmišljati tudi o izboru primerjalnega rezultata, na podlagi katerega se bomo odločili, ali naš model deluje dobro ali slabo. Sama metrika uspešnosti postane informativna šele, ko jo primerjamo z ustreznim merilom. Tako kot pri športnih aktivnostih uspeh atletov merimo na podlagi primerjanja z rezultati ostalih tekmovalcev, mo-

ramo tudi predlagani model primerjati z rezultati ostalih modelov [12].

V nadaljevanju podajamo opis izbranih metrik uspešnosti, s katerimi bomo izmerili kakovost predlagane trgovalne strategije in napovednega modela. Ključno za naš uspeh je, da tako trgovalna strategija kot tudi napovedni model delujeta dobro. Dober napovedni model z zelo slabo trgovalno strategijo se bo sicer dobro odrezal pri nekaterih metrikah, zelo slab rezultat pa bo dajal pri samem trgovanju.

4.1 Napovedni model

Predlagani napovedni model, opisan v poglavju 3, kot vhod sprejme nabor zgodovinskih *OHLC* podatkov, jih obdela ter rezultira v seznamu napovedi. Vektor napovedi nam pove, ali bo vrednost delnice v bližnji prihodnosti zrastle, padla ali pa bo stagnirala. Če primerjamo naš model s [7], nismo napovedovali vrednosti delnic v prihodnosti, temveč splošen trend gibanja. Če odmislimo domeno našega napovednega modela, lahko napovedovanje posplošimo na klasifikacijo primerov v enega izmed treh ustreznih razredov.

Ker testiramo klasifikacijski model, temu primerno lahko tudi izberemo nabor metrik uspešnosti. Pri problemu klasifikacije pridejo v poštev metrike: klasifikacijska točnost, AUC, logaritmična izguba itd. [17]. Ker naš model rešuje problem večrazredne klasifikacije, mera AUC [18] ne pride v poštev, saj je ta v splošnem namenjena binarni klasifikaciji. V nadaljevanju podajamo opis izbranih metrik ter predstavimo nekaj rezultatov, ki bodo služili kot osnova pri interpretaciji kakovosti predlaganega napovednega modela.

4.1.1 Klasifikacijska točnost

Klasifikacijska točnost je relativno preprosta mera, ki nam pove zgolj koliko odstotkov primerov smo pravilno razvrstili v pripadajoči razred. Formula, po kateri izračunamo klasifikacijsko točnost, je podana v 4.1.

$$accuracy(y, p) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(p_i = y_i) \quad (4.1)$$

V izseku kode 4.1 si lahko na preprostem primeru pogledamo, kakšen rezultat dobimo za podani primer pravih vrednosti in napovedanih razredov.

```
1 y_truth = [0,0,0,2,0,2,2,1,1,1]
2 predictions = [0,2,1,0,0,2,2,2,1,0]
3
4 print("Accuracy: %.2f" % accuracy(y_truth, predictions))
5
6 > Accuracy: 0.50
```

Koda 4.1: Enostaven primer izračuna klasifikacijske točnosti

Pri nadaljnji interpretaciji velja premisliti, kdaj bomo z rezultatom klasifikacijske točnosti predlaganega modela zadovoljni. Kot omenjeno, napovedni model vsak trgovalni dan razporedi v enega izmed treh razredov. Naključna razporeditev primerov v razrede bi nam torej podala rezultat klasifikacijske točnosti okoli vrednosti 33 %. Torej bo predlagani model dobro deloval šele takrat, ko bo uspešno klasificiral več kot 33 % primerov. Seveda moramo upoštevati, da je klasifikacijska točnost lahko tudi zelo visoka pri zelo slabem klasifikacijskem algoritmu. Namreč v primeru podatkov, kjer imamo zelo malo pripadnikov določenga razreda, lahko dosežemo zelo visoko točnost, a zelo majhno natančnost. Natančnost izračunamo tako, da število pravilno klasificiranih primerov (T_p) delimo s številom vseh klasificiranih primerov za ta razred ($T_p + F_p$). Formula za izračun natančnosti za posamezen razred je podana v Formuli 4.2. Pri analizi rezultatov smo zato podrobneje analizirali distribucijo razredov v naših učnih, testnih in validacijskih množicah.

$$P = \frac{Tp}{Tp + Fp} \quad (4.2)$$

4.1.2 Logaritmična izguba

Predstavimo še malo drugačno mero – logaritmično izgubo, ki močno kaznuje napačne klasifikacije ter nagradi konzervativne napovedi [17]. Metrika pri oceni, za razliko od klasifikacijske točnosti, upošteva stopnjo prepričanja, s katero je bil razred določen. Tako v primeru, da naš model s 100 % prepričanjem za samo en primer napove pripadnost razredu 1, ko ta v resnici pripada razredu 2, dobimo neskončen rezultat in tako praktično zavržemo napovedni model kot popolnoma nekoristen. Definicija metrike za binarno klasifikacijo je podana v formuli 4.3. V postopku razvoja smo uporabili programski paket *sklearn*, kjer lahko logaritmično izgubo izračunamo tudi za večrazredne klasifikatorje.

$$\text{logloss}(y, p) = -\frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} (y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \quad (4.3)$$

Za razliko od klasifikacijske točnosti, kjer stremimo k rezultatu 1, je najboljši možen rezultat pri logaritmični izgubi 0,0. Seveda takšnega rezultata napovedni model ne bo nikoli dosegel, saj bi moral v tem primeru prav vse primere s 100 % prepričanjem razvrstiti v pravilni razred. Poglejmo si torej, kakšen rezultat dobimo v primeru, ko naš napovedni model za vsakega izmed primerov napove enako verjetno pripadnost kateremukoli izmed treh razredov v rezultatu 4.4.

$$\text{logloss}(\text{uniform}) = 1,10 \quad (4.4)$$

V postopku analize rezultatov bomo tako torej stremeli k uporabi parametrov modela, ki bodo dosegali logaritmične izgube, nižje od vrednosti 1,10.

4.2 Trgovalna strategija

Ker želimo testirati predlagani napovedni model tudi s samim trgovanjem delnic oziroma s simulacijo le-tega, smo morali razviti tudi trgovalno strategijo, na podlagi katere izvajamo dejansko simulacijo. Napovedni model smo tako testirali tudi s simulacijo trgovanja, kjer smo uspeh modela merili v višini donosa v daljšem časovnem obdobju.

Preden se podamo v samo testiranje napovednega modela, se moramo prepričati o kakovosti predlagane trgovalne strategije. Rezultate testiranja trgovalne strategije podamo v poglavju 5.1 ter 5.2, kjer preverimo samo odpornost strategije na napake. Naša osnovna hipoteza pri trgovalni strategiji je, da nam njena kakovost omejuje tudi kakovost napovednega modela. Torej na nek način s slabo trgovalno strategijo predefiniramo tudi delovanje napovednega modela, ki seveda v simulaciji trgovanja ne bo prinašal visokih donosov. Nasprotno nam dobro zastavljena trgovalna strategija z dobro definiranimi pogoji nakupa in prodaje delnic, pri katerih ni upoštevan zgolj napovedni model, temveč tudi ostali rezultati tehnične analize, lahko prikaže boljše delovanje, kot bi ga modelu smeli pripisati.

Naš cilj ni maksimizirati dobiček oziroma minimizirati izgubo z odlično trgovalno strategijo, temveč postaviti ogrodje, s katerim lahko objektivno ocenimo kakovost predlaganega napovednega modela. Zato predlagana trgovalna strategija ne vsebuje dodatnih izračunov tehnične analize, temveč deluje zgolj na podlagi predlaganega napovednega modela.

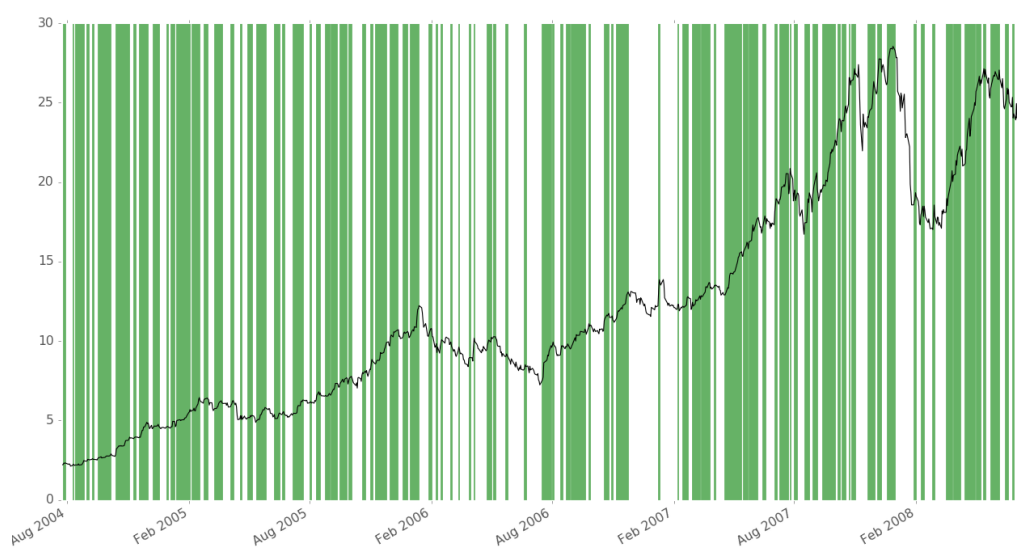
4.2.1 Simulator

Trgovanje na podlagi napovednega modela smo opravili v okviru simulacije na podlagi zgodovinskih podatkov. Naloga razvitega simulatorja je v prvi vrsti testiranje kakovosti delovanja napovednega modela kot tudi same trgovalne strategije. Želeni parametri simulacije so:

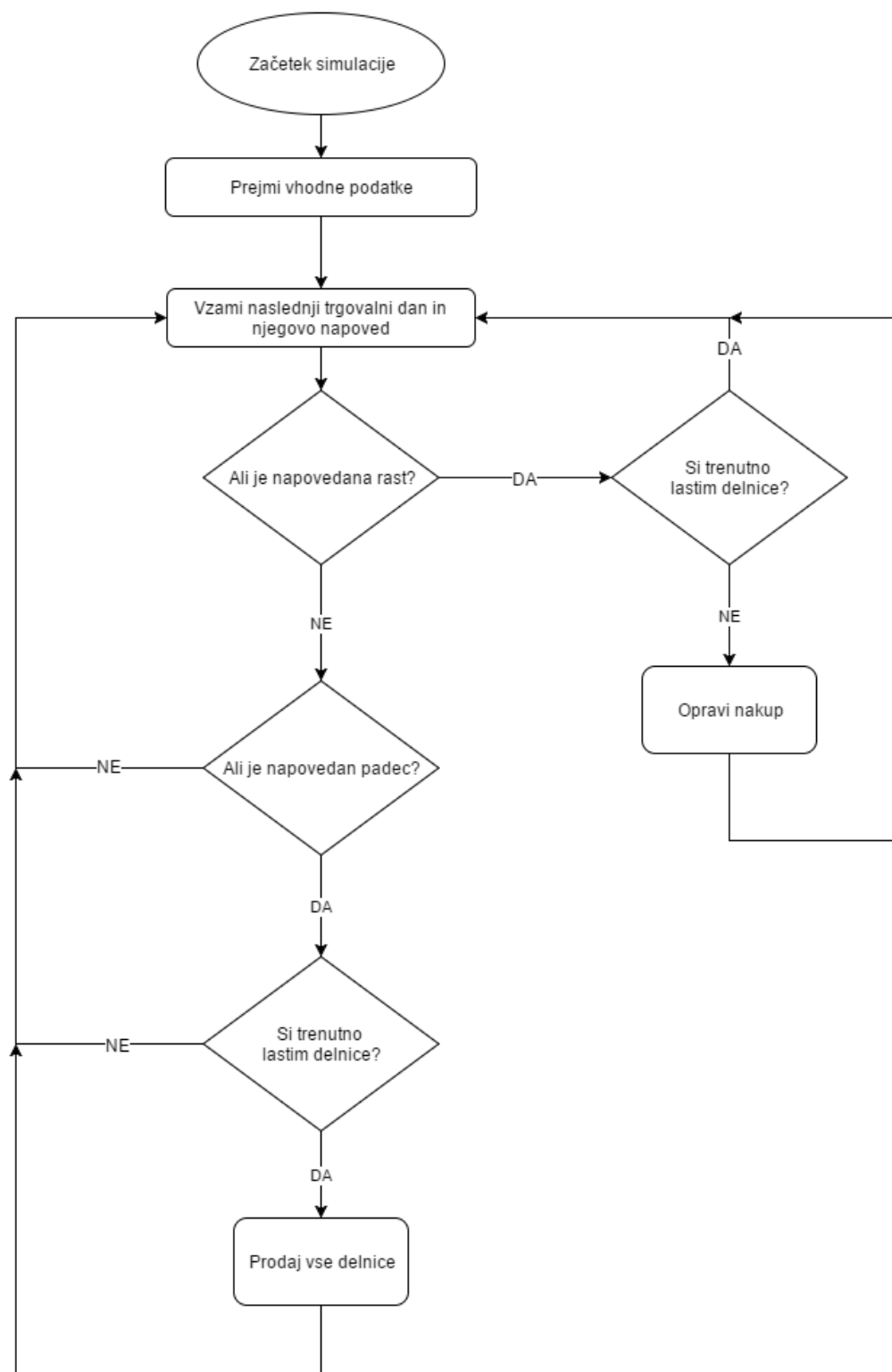
- osnovni kapital,
- obdobje simulacije,
- višina trgovalne provizije,
- izbrana delnica,
- trgovalna strategija.

Na izhodu pričakujemo končno stanje kapitala in število prodaj in nakupov. Slednje je poleg končnega stanja zelo pomembno, saj lahko s pomočjo tega izračunamo, koliko denarja smo namenili za pokritje provizij. Tako recimo tudi model, ki je morda končal z nizkim zneskom kapitala, a veliko trgovalnimi akcijami, ni popolnoma napačen, le trgovalne provizije ni najbolje prenesel. Seveda si želimo, da bi predlagani model imel visok donos in majhno število trgovalnih akcij. Dodatno si želimo kot izhod simulacije pridobiti še graf gibanja vrednosti delnic ter označena obdobja, ko je predlagani napovedni model delnice kupil ter prodal. Primer takšnega grafa podajamo na Sliki 4.1. Iz takšnega grafa lahko razberemo, ali je napovedni model prožil signale za nakup v obdobjih rasti ter ali je pravočasno prožil signale za prodajo pred padci. Podrobnejše delovanje simulatorja je prikazano na Sliki 4.2.

Primerjava rezultatov in njihova ocena kakovosti je precej preprosta. Če smo obdobje simulacije pričeli s \$1.000 si želimo obdobje zaključiti z zneskom, ki bo višji od začetnega stanja. Osnovna model, ki bo služil za primerjavo s predlaganim napovednim modelom, je model *Kupi in zadrži*, katere podrobnejši opis podajamo v poglavju 5.1.1. Dodatno bomo kakovost primerjali z modelom *Tekočih povprečij* ter *MACD*.



Slika 4.1: Primer grafa z obdobji lastništva delnic



Slika 4.2: Diagram delovanja simulatorja

Poglavje 5

Analiza rezultatov

Pravilna obdelava podatkov in interpretacija pridobljenih rezultatov je zelo pomembna. Ker obdelujemo časovno odvisne podatke, je pomembno, da v nobeni točki ne upoštevamo informacij ali znanja, ki postane na voljo šele v prihodnosti. Zato smo pri analizi delovanja predlaganega modela namenili dodatno pozornost zadostitvi temu pogoju. OHLC podatke delnic smo v času razvoja razdelili na tri ključne skupine:

1. učno množico,
2. testno množico,
3. validacijsko množico.

Učne množice vsebujejo prvih 1.000 OHLC podatkov, testne naslednjih 1.000 ter validacijske zadnjih 2.000 za posamezno delnico. Pomembno je poudariti, da med samim razvojem modela nismo nikoli pregledali ali analizirali validacijske množice, ker bi s tem lahko zavestno ali podzavestno prilagodili parametre ali metode za optimalni rezultat – pristranskost vpogleda v prihodnost [12]. V obdobju razvoja smo analizirali zgolj učno ter testno množico. Razvite algoritme in modele smo testirali, kot tudi samo ime nakazuje, na testni množici, kjer smo iskali optimalne parametre za optimalni rezultat. Model z naborom parametrov, ki so se dobro odrezali na testni množici, smo

nato pogнали še na validacijski množici podatkov. V nadaljevanju podajamo analizo delovanja trgovalne strategije in razvitih modelov za izbrane delnice.

5.1 Testiranje trgovalne strategije

V poglavju 4.2.1 smo opisali delovanje simulatorja, s katerim želimo testirati kakovost trgovalne strategije, opisane v poglavju 3.5.2, ter seveda tudi kakovost napovednega modela. Če na kratko povzamemo trgovalno strategijo, želimo prožiti signale nakupa delnice takrat, ko njena vrednost v prihajajočih dneh preseže 2-kratnik trgovalne provizije, opisane v poglavju 3.5.1. Podobno velja za signal za prodajo – prožimo ga takrat, ko zaznamo, da bo delnica v prihajajočih dneh padla pod 2-kratnik trgovalne provizije.

Poenostavljen grafični prikaz trgovalne strategije smo prikazali na Sliki 3.6. Parametri, ki določajo našo trgovalno strategijo, so:

1. koliko dni v prihodnost gledamo,
2. višina osnovnega kapitala,
3. višina trgovalne provizije.

Višino provizije smo določili v poglavju 3.5.1, njeno vrednost pa smo nastavili na \$15.

Parameter *period_frame_days* predstavlja število dni, kolikor se z drsečim oknom pri vsakem OHLC podatku premaknemo naprej – s pomočjo tega označimo posamezne OHLC podatke z ustreznim razredom:

- **Razred 0** – vrednost delnice v prihodnjih x dneh preseže 2-kratnik trgovalne provizije,
- **Razred 1** – vrednost delnice v prihodnjih x dneh pade pod 2-kratnik trgovalne provizije,
- **Razred 2** – vrednost delnice v prihodnjih x dneh se ne premakne ne čez zgornjo niti pod spodnjo mejo.

Z zgornjim opisom postane hitro razvidno, da več dni kot bomo gledali v prihodnost, manj bomo imeli predstavnikov razreda **2**, saj bo delnica eventualno zrastle ali padla nad oziroma pod določeno mejo.

5.1.1 Kupi in zadrži

Predlagano trgovalno strategijo najprej preizkusimo z najbolj osnovnim napovednim modelom oziroma strategijo *Kupi in zadrži* (angl. *Buy and Hold*), katere veliki zagovornik je tudi priznan in zelo uspešen investitor g. Warren Buffet.¹

Strategija *Kupi in zadrži* je zelo preprosta. Ob začetku trgovanja delnice kupimo in jih zadržimo do zadnjega dneva, ko jih prodamo oziroma izračunamo njihovo vrednost na zadnji trgovalni dan. Naša sredstva se v tem primeru gibljejo popolnoma enako kot vrednost delnice, donos pa je tudi enak rasti delnice čez neko daljše obdobje. Omenjeno strategijo smo uporabili za primerjavo in merilo delovanja predlaganega napovednega modela [12]. Naš cilj je, da razviti model v okviru trgovalne strategije deluje bolje od *Kupi in zadrži*. Rezultati donosov strategije *Kupi in zadrži*, pri začetnem vložku v višini \$1.000, so za izbrane delnice v testnem kot tudi v validacijskem obdobju podane v Tabeli 5.1.

| Delnica | Testno obdobje | Validacijsko obdobje |
|-----------|----------------|----------------------|
| Coca-Cola | \$199,68 | \$513,94 |
| Microsoft | - \$292,70 | \$1.918,17 |
| Apple | \$10.255,71 | \$2.983,09 |

Tabela 5.1: Donosi za Kupi in zadrži

Razberemo lahko, da vse delnice v validacijskem obdobju rastejo, saj s strategijo *Kupi in zadrži* vse tri delnice v tem obdobju ustvarijo dobiček. Seveda se med seboj izbrane delnice precej razlikujejo po sami stopnji rasti,

¹<http://www.cnbc.com/2016/03/04/warren-buffett-buy-hold-and-dont-watch-too-closely.html>

tako npr. Microsoftove v testnem obdobju med prvim in zadnjim trgovalnim dnevom celo izgubijo vrednost, saj iz začetnih \$1.000 v prvem dnevu zadnji trgovalni dan zaključimo z zgolj \$707,30 kapitala, kar pomeni \$292,70 izgube.

5.1.2 Izbira parametrov

Izbira ustreznih vrednosti parametrov ima pri napovedovanju zelo pomembno vlogo. Med pomembnejšimi parametri smo že izpostavili parameter *period_frame_days*, s katerim določimo kako daleč v prihodnost gledamo, ko označujemo podatke delnic z razredi **0**, **1** ali **2**. Parameter na neki način omejuje število transakcij, ki jih naša trgovalna strategija izvede, saj v primeru razreda **2**, ko gre za stagnacijo, naša trgovalna strategija ne bo izvedla trgovalne akcije. Vsekakor je optimalna vrednost omenjenega parametra odvisna od višine trgovalne provizije in osnovnega kapitala, s katerima določimo, kakšni sta meji, prikazani na Sliki 3.6, ki prožita signale za nakup oziroma prodajo.

Rezultati v Tabeli 5.2 prikazujejo, kako se predlagana trgovalna strategija obnese, če bi napovedni model znal zagotovo pravilno napovedati razred posameznega dneva – tak napovedni model poimenujemo *FutureTeller*. *FutureTeller* se za vsak podani *OHLC* podatek sprehodi toliko dni v prihodnost, kot to določa parameter *period_frame_days*, in določi pravilen razred podanega dneva – **0**, **1** ali **2**.

Kot lahko razberemo iz Tabele 5.2, je optimalna vrednost parametra *period_frame_days* za različne delnice različna. V primeru Coca-Cole vidimo, da podana strategija proži najvišje donose pri vrednosti parametra 10 – torej ko se za vsako napoved sprehodimo kar 10 dni v prihodnost. Če bi naš napovedni model znal pravilno napovedati pripadajoči razred za 10 dni vnaprej, bi tako pri delnici Coca-Cole v testnem obdobju iz začetnih \$1.000 pridelali dobrih \$1.240,60 dobička. V primeru Microsoftove delnice lahko vidimo, da bi naša strategija rezultirala v najvišjih donosih pri 8 dneh vpogleda v prihodnost. Najvišje donose pri delnici podjetja Apple pa bi pridelali, če bi znali napovedati spremembo delnice za zgolj 3 dni vnaprej – kar neverje-

| Vrednost parametra | Apple | Coca-Cola | Microsoft |
|--------------------|-----------------------------|------------------------|--------------------------|
| 1 | \$218.634,33 (84) | \$444,72 (2) | \$4.045,17 (49) |
| 2 | \$1.331.975,57 (135) | \$358,72 (5) | \$10.775,08 (77) |
| 3* | \$1.951.681,48 (154) | \$569,78 (11) | \$18.239,32 (91) |
| 4 | \$1.914.107,33 (160) | \$911,38 (21) | \$19.662,83 (95) |
| 5 | \$1.675.228,07 (162) | \$1.109,46 (29) | \$19.816,59 (103) |
| 6 | \$1.570.702,26 (172) | \$1.121,78 (31) | \$19.628,19 (109) |
| 7 | \$1.501.778,75 (178) | \$1.197,48 (36) | \$21.982,26 (117) |
| 8* | \$1.329.938,22 (182) | \$1.158,62 (40) | \$22.332,96 (119) |
| 9 | \$1.389.022,92 (194) | \$1.222,43 (42) | \$20.223,74 (123) |
| 10* | \$1.349.404,02 (196) | \$1.240,60 (44) | \$19.887,14 (127) |
| 11 | \$1.336.948,19 (200) | \$1.230,59 (48) | \$20.535,25 (127) |
| 12 | \$1.288.025,88 (200) | \$1.162,42 (50) | \$19.858,71 (131) |
| 13 | \$1.279.826,58 (202) | \$1.125,43 (54) | \$18.640,23 (131) |
| 14 | \$1.248.548,85 (204) | \$966,38 (58) | \$18.958,74 (131) |
| 15 | \$1.257.288,40 (204) | \$847,53 (58) | \$19.159,28 (133) |
| 16 | \$1.256.333,48 (204) | \$852,34 (60) | \$17.863,73 (133) |
| 17 | \$1.264.185,53 (206) | \$775,28 (60) | \$17.945,06 (133) |
| 18 | \$1.274.271,32 (208) | \$762,66 (60) | \$17.360,02 (133) |
| 19 | \$1.274.271,32 (208) | \$684,47 (62) | \$15.451,64 (135) |
| 20 | \$1.289.545,85 (210) | \$568,81 (64) | \$14.362,88 (135) |
| 21 | \$1.289.545,85 (210) | \$509,36 (66) | \$13.721,56 (135) |
| 22 | \$1.289.545,85 (210) | \$493,92 (66) | \$13.608,18 (135) |
| 23 | \$1.289.545,85 (210) | \$394,23 (68) | \$12.488,46 (137) |
| 24 | \$1.289.545,85 (210) | \$362,77 (68) | \$11.838,48 (137) |
| 25 | \$1.289.545,85 (210) | \$301,33 (68) | \$11.421,56 (137) |
| 26 | \$1.289.545,85 (210) | \$256,49 (68) | \$10.880,14 (137) |
| 27 | \$1.289.545,85 (210) | \$141,77 (70) | \$9.981,20 (137) |
| 28 | \$1.289.545,85 (210) | \$142,17 (70) | \$9.372,74 (139) |
| 29 | \$1.289.545,85 (210) | \$112,99 (72) | \$8.219,81 (141) |
| 30 | \$1.289.545,85 (210) | \$34,58 (72) | \$6.820,79 (143) |
| 31 | \$1.289.545,85 (210) | \$34,84 (72) | \$6.913,37 (143) |
| 32 | \$1.289.545,85 (210) | \$33,76 (72) | \$6.819,19 (143) |
| 33 | \$1.289.545,85 (210) | \$11,04 (72) | \$6.491,24 (143) |
| 34 | \$1.289.545,85 (210) | - \$41,19 (72) | \$6.514,65 (143) |
| 35 | \$1.289.545,85 (210) | - \$172,17 (74) | \$6.514,65 (143) |

Tabela 5.2: Donos in število trgovalnih akcij (v oklepajih), za metodo FutureTeller pri različnih vrednostih parametra *period_frame_days*

tnih \$1.951.681,48 dobička bi pridelali iz skromnih \$1.000 začetne investicije. Vrednosti v oklepajih predstavljajo število trgovalnih akcij.

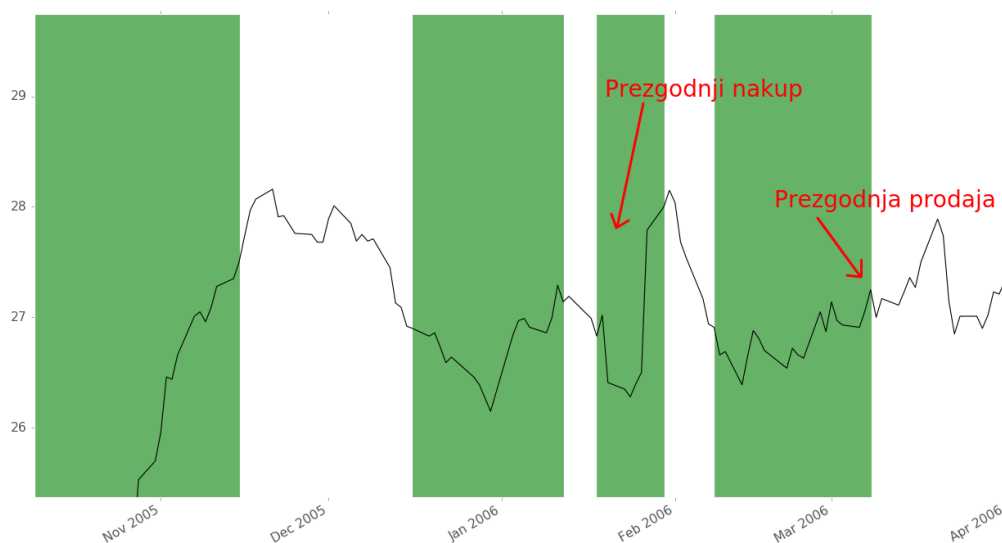
Seveda moramo rezultate *FutureTeller* upoštevati z nekaj zadržki, saj bi takšne donose videli zgolj v primeru, če bi napovedni model vedno in brez izjeme vrnil pravilno napoved. Takšne rezultate je povsem nemogoče ali pa zelo težko pričakovati.

Velja razmisliti še, zakaj se donosi kljub popolnemu napovednemu modelu pri večjih vrednostih parametra *period_frame_days* zmanjšujejo. Morda se zdi na prvi pogled bolj intuitivno, da dlje kot gledamo v prihodnost, boljši bo rezultat. Seveda ne gre oporekati, da so rezultati tudi v teh primerih precej dobri, a najboljše rezultate dobimo pri relativno kratkih vpogledih v prihodnost. Upoštevati je treba delovanje predlagane trgovalne strategije – vedno se odločamo, ali bomo delnice kupili ali prodali danes. Pri krajših vpogledih v prihodnost dopuščamo večjo verjetnost pojavitve razreda **2** – v prihodnjih nekaj dneh se vrednost ne spremeni dovolj, da bi opravičila nakup oziroma prodajo. Tako torej lahko sklepamo, kar potrjujejo tudi rezultati, da dlje kot gledamo v prihodnost, manj imamo predstavnikov razreda **2** in vedno več predstavnikov razredov **0** ali **1** – le da so te spremembe lahko precej daleč v prihodnosti. S tem povzročimo dve neželeni posledici:

1. povečamo število trgovalnih akcij,
2. prehitro prožimo signale za nakup oziroma prodajo.

Primer prehitro proženih signalov za nakup prikazuje Slika 5.1. V zadnjem zelenem pasu, ki označuje obdobje, ko si lastimo delnico, lahko vidimo primer prehitro proženega signala za prodajo – vrednost delnice je nekaj dni po proženem signalu še vedno rastla. Podobno lahko na predzadnjem odseku vidimo, da smo prožili signal za nakup, še preden je delnica dosegla lokalni minimum.

Podatki v Tabeli 5.2 temeljijo na izračunih, kjer smo meje za proženje nakupnih in prodajnih signalov postavili na osnovi \$1.000 začetnega kapitala in trgovalni proviziji \$15. Ker to pomeni, da mora delnica v naslednjih



Slika 5.1: Primer prezgodaj in prepozno proženih signalov

n dneh zrasti ali pasti za dobre 3 %, je število proženih signalov relativno majhno. Poglejmo si še, kakšne rezultate bi dobili v primeru, kjer za parameter začetnega kapitala izberemo vrednost \$5.000, kjer signale prožimo precej bolj pogosto, saj mora vrednost delnice zrasti ali pasti le še za 0,6 %. Iz rezultatov v Tabeli 5.3 lahko potrdimo našo hipotezo, da bo število trgovalnih akcij precej višje kakor poprej. Model sedaj proži akcijo praktično vsak drugi ali tretji dan – vseh trgovalnih dni imamo v testnem obdobju zgolj 1.000. Opazimo tudi, da je število dni, potrebnih za vpogled v prihodnost, manjše kakor poprej – zgolj 3 ali 4 dni. *FutureTeller* je zdaj sposoben praktično zaznati skoraj vsak porast ali padec delnic, kar se kaže v večjem številu transakcij in tudi občutno večjem donosu. Potrditev, da *FutureTeller* s takšnimi parametri zares zazna večino lokalnih maksimumov in minimumov, lahko razberemo tudi s Slike 5.2.

| Vrednost parametra | Apple | | Coca-Cola | | Microsoft | |
|--------------------|------------------|-------|-------------|-------|----------------|-------|
| 1 | \$113.877.890,39 | (387) | \$43.892,83 | (213) | \$1.104.266,75 | (299) |
| 2 | \$121.838.628,02 | (389) | \$58.246,07 | (254) | \$1.333.974,01 | (325) |
| 3 | \$121.651.641,37 | (389) | \$58.325,13 | (264) | \$1.337.120,04 | (327) |
| 4 | \$121.880.347,50 | (393) | \$57.371,22 | (268) | \$1.340.662,24 | (327) |
| 5 | \$121.794.582,17 | (393) | \$54.902,08 | (274) | \$1.292.508,13 | (329) |
| 6 | \$121.419.871,70 | (393) | \$54.054,16 | (274) | \$1.269.333,22 | (329) |
| 7 | \$121.419.871,70 | (393) | \$53.284,06 | (282) | \$1.273.597,24 | (329) |
| 8 | \$121.419.871,70 | (393) | \$52.689,03 | (282) | \$1.273.597,24 | (329) |
| 9 | \$121.419.871,70 | (393) | \$52.744,62 | (284) | \$1.273.597,24 | (329) |
| 10 | \$121.419.871,70 | (393) | \$52.887,67 | (284) | \$1.273.597,24 | (329) |
| 11 | \$121.419.871,70 | (393) | \$52.659,53 | (284) | \$1.273.597,24 | (329) |
| 12 | \$121.419.871,70 | (393) | \$52.659,53 | (284) | \$1.273.597,24 | (329) |
| 13 | \$121.419.871,70 | (393) | \$52.659,53 | (284) | \$1.273.597,24 | (329) |
| 14 | \$121.419.871,70 | (393) | \$52.659,53 | (284) | \$1.273.597,24 | (329) |
| 15 | \$121.419.871,70 | (393) | \$52.659,53 | (284) | \$1.273.597,24 | (329) |
| 16 | \$121.419.871,70 | (393) | \$52.659,53 | (284) | \$1.273.597,24 | (329) |
| 17 | \$121.419.871,70 | (393) | \$52.659,53 | (284) | \$1.273.597,24 | (329) |
| 18 | \$121.419.871,70 | (393) | \$52.659,53 | (284) | \$1.273.597,24 | (329) |
| 19 | \$121.419.871,70 | (393) | \$52.659,53 | (284) | \$1.273.597,24 | (329) |
| 20 | \$121.419.871,70 | (393) | \$52.659,53 | (284) | \$1.273.597,24 | (329) |

Tabela 5.3: Donos in število trgovalnih akcij (v oklepajih), za metodo Future-Teller pri različnih vrednostih parametra *period_frame_days* - osnovni kapital \$5.000



Slika 5.2: Graf delovanja metode FutureTeller pri \$5.000 osnovnega kapitala

| pf.days | 100% | 95% | 90% | 85% | 80% |
|---------|-----------------|-----------------|--------------------|--------------------|--------------------|
| 1 | \$444,72 (2) | - \$366,01 (28) | - \$832,05 (57) | - \$1.012,10 (85) | - \$1.007,30 (73) |
| 2 | \$358,72 (5) | - \$70,20 (26) | - \$765,29 (55) | - \$997,03 (84) | - \$988,20 (113) |
| 3 | \$569,78 (11) | - \$23,75 (39) | - \$842,11 (68) | - \$998,73 (96) | - \$988,32 (116) |
| 4 | \$911,38 (21) | \$107,24 (51) | - \$288,37 (67) | - \$989,09 (112) | - \$996,21 (133) |
| 5 | \$1.109,46 (29) | \$142,60 (64) | - \$517,99 (78) | - \$996,33 (117) | - \$1.012,15 (123) |
| 6 | \$1.121,78 (31) | \$475,45 (52) | - \$793,47 (89) | - \$1.012,85 (119) | - \$1.008,96 (159) |
| 7 | \$1.197,48 (36) | \$486,52 (56) | - \$336,69 (78) | - \$1.001,39 (128) | - \$1.006,46 (148) |
| 8 | \$1.158,62 (40) | \$583,31 (54) | - \$973,35 (98) | - \$997,79 (139) | - \$1.012,83 (176) |
| 9 | \$1.222,43 (42) | \$15,40 (78) | - \$994,72 (112) | - \$994,93 (161) | - \$994,27 (188) |
| 10 | \$1.240,60 (44) | \$98,19 (76) | - \$982,08 (96) | - \$1.013,84 (169) | - \$991,91 (183) |
| 11 | \$1.230,59 (48) | - \$178,36 (80) | - \$663,33 (98) | - \$1.001,71 (148) | - \$1.010,02 (219) |
| 12 | \$1.162,42 (50) | \$138,63 (70) | - \$989,75 (127) | - \$1.010,95 (192) | - \$1.002,70 (205) |
| 13 | \$1.125,43 (54) | - \$285,03 (84) | - \$995,03 (152) | - \$1.005,67 (200) | - \$989,52 (242) |
| 14 | \$966,38 (58) | - \$263,12 (82) | - \$1.001,78 (114) | - \$992,07 (143) | - \$988,57 (266) |
| 15 | \$847,53 (58) | - \$483,68 (92) | - \$958,56 (102) | - \$1.001,73 (249) | - \$989,28 (260) |

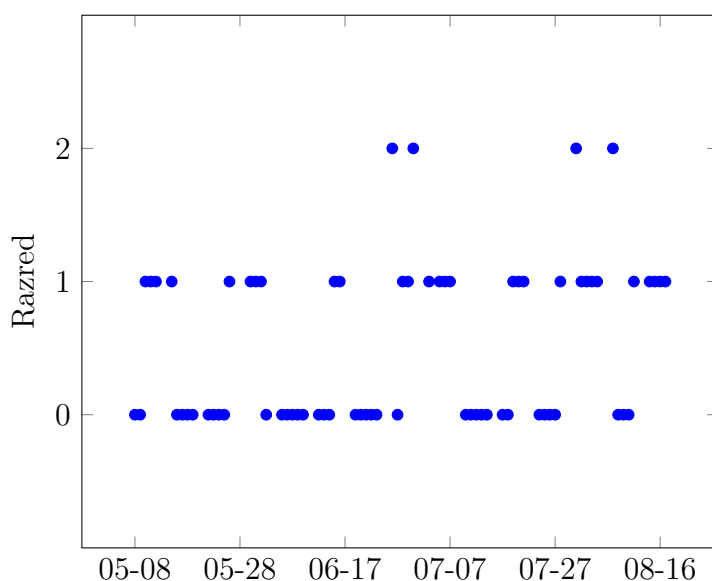
Tabela 5.4: Odpornost trgovalne strategije na napake na podatkih delnice Coca-Cola

5.2 Kakovost trgovalne strategije

V prejšnjem poglavju smo spoznali vpliv parametrov na našo trgovalno strategijo. V Tabeli 5.2 smo podali rezultate za primer napovednega modela *FutureTeller*, ki vedno napove pravilen razred. Kot smo že omenili, je takšen napovedni model povsem nerealno pričakovati, zato so rezultati v Tabeli 5.2 zgolj informacija, ali naša strategija deluje dobro v idealnih pogojih. Ker predlagani, pa tudi katerikoli drugi napovedni model nikakor ne bo mogel doseči 100 % pravilne napovedi, je smiselno trgovalno strategijo testirati tudi na manj ugodnih pogojih, kjer želimo preveriti robustnost in odpornost trgovalne strategije na napake.

V Tabeli 5.4 so podani rezultati za primere, ko napovedni model pravilno napove verjetnost razreda med 80 in 100 % primerov. Napovedni model je bil testiran na delnicah Coca-Cole pri višini osnovnega kapitala \$1.000. Vidimo, da so pričakovani donosi drastično upadli, saj šele pri klasifikacijski točnosti 95 % postanejo le-ti pozitivni, pri nižjih točnostih pa trgovalna strategija žal ne prinaša dobičkov. To je seveda precej slaba novica, saj pomeni, da predlagana trgovalna strategija ni odporna na napake ali kakršenkoli šum v napovednem modelu.

Ker predlagani napovedni model, najverjetneje pa tudi kakšen drug, ne bo dosegel tako visoke klasifikacijske točnosti, je treba poiskati vzrok za



Slika 5.3: Sosledje razredov

tako slabo delovanje in ga poskušati odpraviti. Najprej si pogledjmo, kako si časovno sledijo pravilni razredi v učni množici na Sliki 5.3. Kot lahko vidimo, si podatki praktično sledijo v vrsti – torej ko pride do obdobja padca, imamo zaporednih več pripadnikov razreda **1**, enako imamo pri obdobju rasti v vrsti več zaporednih primerov razreda **0**. To informacijo uporabimo pri izboljšavi simulatorja oziroma trgovalne strategije. V postopek simulacije podamo dodatno zahtevo, da mora napovedni model vsaj 2 dni zapored prožiti signal za nakup, da delnice dejansko kupimo. Podobno velja za signal za prodajo – signal moramo prožiti 2-krat zapored, da delnice dejansko prodamo.

Poglejmo si rezultate posodobljene trgovalne strategije v Tabeli 5.5. Rezultati so tokrat malenkost boljši – strategija deluje dobro tudi pri slabših napovednih modelih. Seveda smo z izboljšavo tudi delno znižali zgornjo mejo zaslužka. Če primerjamo rezultate iz Tabele 5.2, lahko razberemo, da pri 100 % pravih napovedi za delnice Coca-Cole pri vrednosti parametra *period_frame_days* 10 sedaj ne zaslužimo več \$1.240,60, temveč \$1.076,93.

| pf_days | 100% | 90% | 80% | 70% | 60% |
|---------|-----------------|-----------------|-----------------|-------------------|--------------------|
| 1 | \$316,10 (1) | - \$415,74 (23) | - \$492,03 (40) | - \$867,00 (60) | - \$1.001,33 (79) |
| 2 | \$608,46 (3) | - \$61,37 (26) | - \$664,67 (39) | - \$782,04 (55) | - \$960,22 (74) |
| 3 | \$456,88 (5) | - \$59,71 (26) | - \$231,12 (41) | - \$853,31 (65) | - \$994,13 (79) |
| 4 | \$772,66 (13) | \$210,84 (34) | - \$528,69 (43) | - \$816,17 (73) | - \$1.004,00 (108) |
| 5 | \$801,37 (17) | \$41,79 (31) | - \$449,37 (52) | - \$988,77 (83) | - \$989,99 (110) |
| 6 | \$915,01 (19) | \$531,45 (35) | - \$305,66 (57) | - \$678,44 (77) | - \$1.002,50 (105) |
| 7 | \$829,08 (23) | \$203,57 (49) | \$261,73 (46) | - \$756,89 (74) | - \$1.000,95 (114) |
| 8 | \$835,88 (28) | \$621,64 (54) | - \$412,50 (61) | - \$588,78 (76) | - \$989,79 (91) |
| 9 | \$1.072,87 (28) | \$419,83 (44) | - \$4,13 (66) | - \$616,88 (78) | - \$869,89 (92) |
| 10 | \$1.076,93 (30) | \$478,66 (52) | \$300,83 (56) | - \$511,13 (78) | - \$985,27 (117) |
| 11 | \$1.165,01 (34) | \$452,93 (46) | - \$123,36 (72) | - \$599,01 (80) | - \$1.010,48 (117) |
| 12 | \$1.034,66 (36) | \$591,82 (46) | - \$445,57 (76) | - \$1.001,76 (99) | - \$1.004,06 (114) |
| 13 | \$1.095,68 (38) | \$419,27 (60) | \$50,66 (72) | - \$628,46 (82) | - \$1.001,72 (122) |
| 14 | \$1.129,71 (40) | \$757,35 (44) | - \$185,71 (74) | - \$719,16 (82) | - \$1.008,30 (120) |
| 15 | \$1.130,37 (40) | \$562,94 (56) | - \$596,79 (82) | - \$608,86 (78) | - \$996,63 (136) |

Tabela 5.5: Odpornost na napake, kjer trgovalna strategija zahteva vsaj dve enaki napovedi na podatkih delnice Coca-Cola

| pf_days | 100% | 90% | 80% | 70% | 60% |
|---------|--------------------|--------------------|--------------------|-------------------|-------------------|
| 1 | \$30.578,07 (36) | \$12.357,10 (57) | \$3.438,39 (69) | \$2.706,25 (76) | \$433,79 (101) |
| 2 | \$144.591,95 (69) | \$107.023,09 (85) | \$38.288,54 (93) | \$13.465,01 (100) | \$4.694,67 (119) |
| 3 | \$259.191,00 (86) | \$132.995,25 (100) | \$95.010,62 (104) | \$44.632,01 (128) | \$13.699,99 (138) |
| 4 | \$422.540,62 (102) | \$233.512,05 (112) | \$108.020,00 (113) | \$76.986,31 (124) | \$33.293,31 (145) |
| 5 | \$484.914,66 (108) | \$299.612,99 (122) | \$191.935,78 (112) | \$46.088,24 (128) | \$24.404,85 (134) |
| 6 | \$435.728,72 (108) | \$323.997,05 (110) | \$152.129,54 (124) | \$83.325,18 (134) | \$34.401,69 (152) |
| 7 | \$426.582,69 (108) | \$199.200,66 (118) | \$203.103,55 (124) | \$58.412,90 (146) | \$15.603,66 (134) |
| 8 | \$435.263,48 (108) | \$210.797,38 (122) | \$195.547,02 (138) | \$79.174,10 (120) | \$23.313,47 (131) |
| 9 | \$442.565,95 (108) | \$288.066,65 (122) | \$137.438,27 (136) | \$74.879,28 (130) | \$18.581,31 (143) |
| 10 | \$418.922,18 (108) | \$239.027,40 (118) | \$103.400,07 (122) | \$42.130,65 (134) | \$18.663,63 (159) |
| 11 | \$420.660,99 (110) | \$265.704,23 (116) | \$102.433,13 (136) | \$73.470,07 (138) | \$16.475,32 (150) |
| 12 | \$403.971,00 (110) | \$241.131,21 (124) | \$182.440,30 (134) | \$64.592,59 (146) | \$39.097,79 (146) |
| 13 | \$407.165,10 (110) | \$262.831,67 (126) | \$173.984,60 (126) | \$70.533,50 (139) | \$11.601,52 (148) |
| 14 | \$408.669,67 (110) | \$128.955,96 (130) | \$83.040,23 (132) | \$57.674,90 (140) | \$8.926,80 (150) |
| 15 | \$394.172,70 (110) | \$213.318,09 (124) | \$119.159,85 (132) | \$26.385,20 (148) | \$30.731,98 (130) |

Tabela 5.6: Odpornost na napake, kjer trgovalna strategija zahteva vsaj dve enaki napovedi na podatkih delnice Apple

Poglejmo si še rezultate odpornosti na napake na delnicah podjetja Apple in Microsoft v tabelah 5.6 in 5.7.

Iz podanih tabel lahko razberemo, da naša trgovalna strategija ni popolna in vsebuje mnogo prostora za izboljšave, a bo za potrebe merjenja kakovosti napovednega modela zadovoljiva. Pri predlaganem napovednem modelu pričakujemo pozitiven donos, glede na samo odpornost trgovalne strategije, pri delnicah podjetja Microsoft ter podjetja Apple, medtem ko za delnice podjetja Coca-Cola ne pričakujemo pozitivnih donosov.

Zgornja analiza trgovalne strategije potrjuje našo začetno trditev, da je v mnogih primerih trgovalna strategija v kontekstu donosa morda celo bolj

| pf_days | 100% | | 90% | | 80% | | 70% | | 60% | |
|---------|------------|------|------------|------|------------|-------|------------|-------|--------------|-------|
| 1 | \$6,67 | (23) | - \$0,35 | (41) | - \$568,39 | (46) | - \$836,21 | (64) | - \$1.006,03 | (93) |
| 2 | \$2.080,25 | (39) | \$302,98 | (49) | - \$204,43 | (63) | - \$174,89 | (74) | - \$1.009,19 | (100) |
| 3 | \$3.514,76 | (49) | \$1.896,99 | (59) | \$1.329,80 | (69) | - \$678,72 | (96) | - \$1.010,20 | (125) |
| 4 | \$4.952,02 | (55) | \$2.943,29 | (69) | \$2.442,04 | (85) | \$524,23 | (91) | - \$905,72 | (109) |
| 5 | \$5.933,43 | (55) | \$4.452,41 | (71) | \$1.070,62 | (81) | \$43,65 | (93) | - \$929,58 | (115) |
| 6 | \$6.457,50 | (59) | \$5.127,22 | (81) | \$331,63 | (91) | - \$635,91 | (111) | - \$776,77 | (103) |
| 7 | \$7.398,41 | (59) | \$5.043,36 | (79) | \$4.616,18 | (87) | \$422,57 | (105) | - \$1.008,12 | (133) |
| 8 | \$6.597,87 | (61) | \$3.565,59 | (85) | \$2.706,55 | (83) | \$776,23 | (97) | \$117,71 | (105) |
| 9 | \$6.935,27 | (69) | \$4.866,80 | (73) | \$2.503,05 | (101) | \$1.143,29 | (97) | - \$1.010,13 | (144) |
| 10 | \$6.737,61 | (69) | \$3.060,49 | (85) | \$1.546,79 | (91) | - \$64,85 | (107) | - \$1.001,35 | (129) |
| 11 | \$7.143,07 | (71) | \$4.758,55 | (77) | \$290,40 | (103) | - \$597,09 | (115) | - \$1.010,91 | (119) |
| 12 | \$7.869,00 | (71) | \$5.703,33 | (79) | \$1.732,45 | (105) | \$72,79 | (119) | - \$983,62 | (124) |
| 13 | \$7.647,65 | (71) | \$4.278,05 | (81) | \$1.985,54 | (111) | - \$687,85 | (119) | - \$1.010,17 | (126) |
| 14 | \$7.704,74 | (73) | \$5.452,36 | (77) | \$1.536,91 | (105) | - \$405,10 | (115) | - \$1.005,07 | (145) |
| 15 | \$7.851,90 | (73) | \$5.101,38 | (85) | \$1.977,82 | (103) | - \$56,34 | (121) | - \$1.004,45 | (128) |

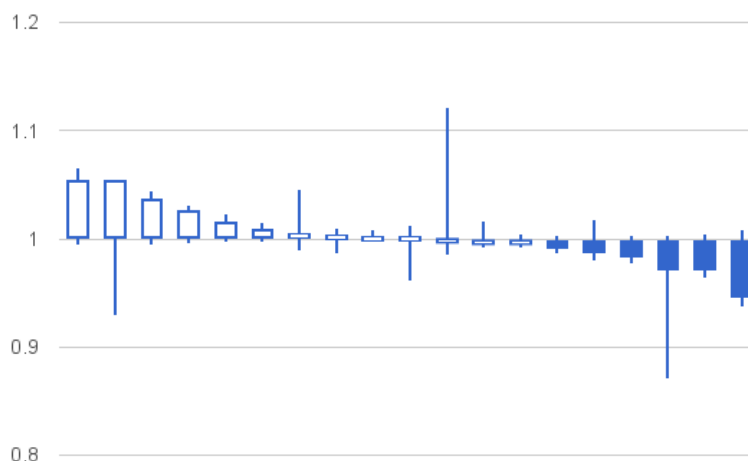
Tabela 5.7: Odpornost na napake, kjer trgovalna strategija zahteva vsaj dve enaki napovedi na podatkih delnice Microsoft

pomembna od napovednega modela. Pokazali smo, da lahko z zelo majhno spremembo drastično povečamo donose pri povsem enako dobrih oziroma slabih napovednih modelih.

5.3 Analiza gruĉ

Japonski sveĉniki so se v zahodnem svetu resno zaĉeli uporabljati davnega leta 1991, ko jih je v svojem delu “Japanese Candlestick Charting Techniques” predstavil Steve Nison [6]. Japonski sveĉniki so vse od takrat naprej precej aktualna tema na podroĉju trgovanja z delnicami – pojavljajo se praktiĉno povsod, kjer so prisotni podatki *OHLC*. V našem delu z japonskimi sveĉniki do zdaj nismo imeli veliko opravka. Kot smo spoznali tudi v poglavju 3, smo japonske sveĉnike pravzaprav iskali in zaznavali povsem avtomatizirano.

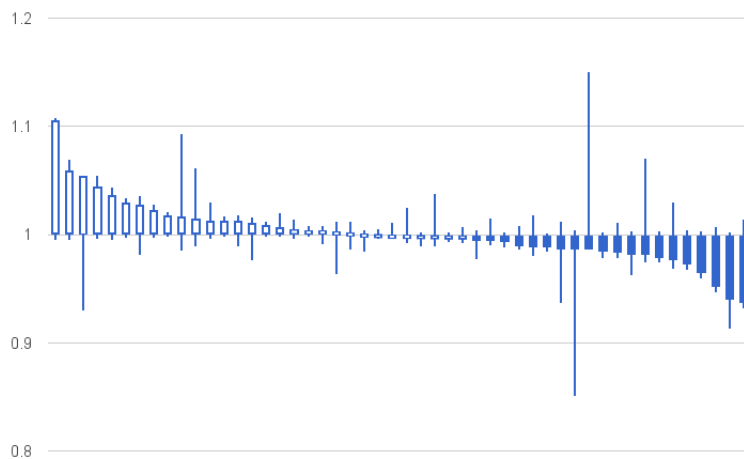
Naša naloga znotraj napovednega modela je bila množico *OHLC* podatkov razvrstiti v skupine japonskih sveĉnikov. To nalogo smo opravili, kot že omenjeno, s pomoĉjo algoritma za gruĉenje *KMeans*. Na Sliki 5.4 so vizualizirani centroidi posameznih gruĉ. Vidimo lahko, da samodejno razvrščanje *OHLC* podatkov rezultira v precej smiselnih gruĉah, saj imamo prisotne tako Dojije, Marubuzute, Spinning Tope in ostale (glej sliko 1.2).



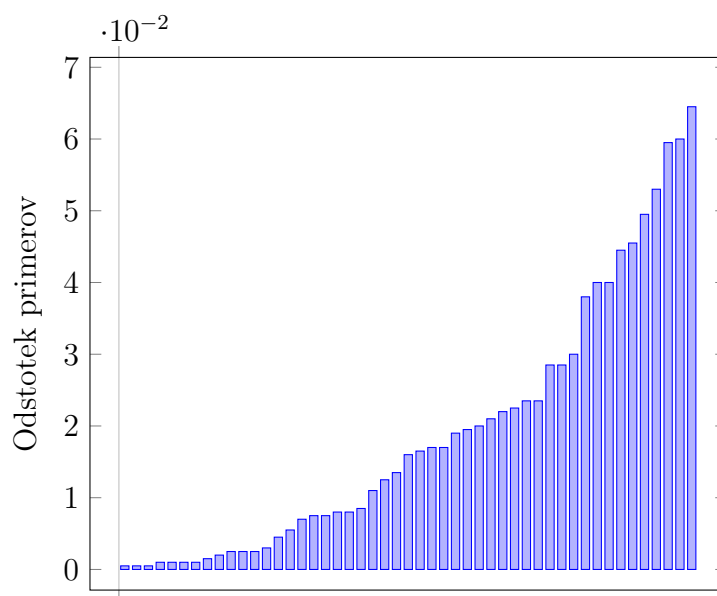
Slika 5.4: Centroidi 20 gruč

Seveda imamo na sliki podanih mnogo več različnih japonskih svečnikov, kot smo jih predstavili v uvodu, saj smo s parametrom algoritma *KMeans* določili, da želimo 20 gruč – skupin japonskih svečnikov. Nekaj centroidov si je med seboj precej podobnih, a se vendarle pomembno razlikujejo tako po sami obliki kot številu pojavitev – tako kot imamo precej besed v naravnem jeziku, ki so si zelo podobne, a imajo povsem različne pomene. Kot primerjavo podajamo še Sliko 5.5, kjer so vizualizirani centroidi v primeru 50 različnih skupin japonskih svečnikov. Tudi tukaj vidimo še vedno precej simiselne skupine, seveda je pričakovano tudi več med seboj podobnih, ki pa se seveda razlikujejo po velikosti telesa ali velikosti senc.

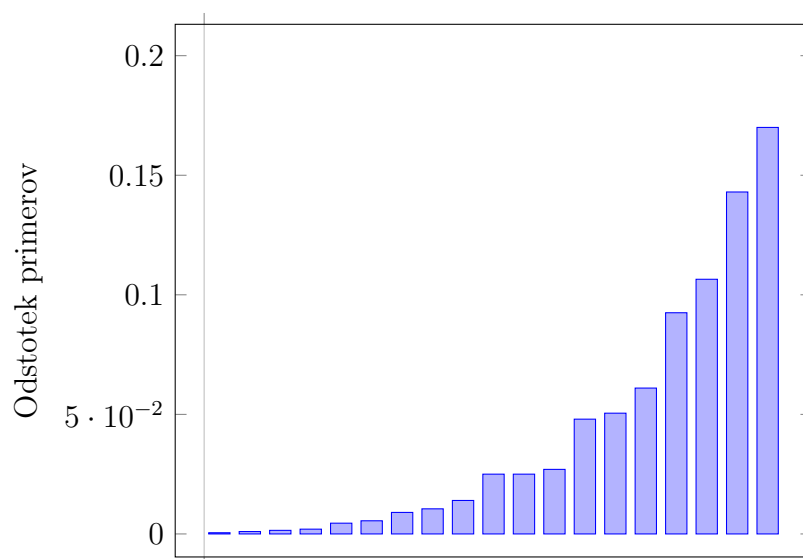
Poglejmo si še porazdelitev *OHLC* podatkov v gruče na Sliki 5.7 in Sliki 5.6. Vidimo lahko, da imamo v obeh primerih nekaj razredov z zelo malo predstavniki ter da je porazdelitev praktično enaka v obeh primerih. Najmanj je seveda robnih primerov – ogromna razlika med začetno in zaključno ceno, bodisi v negativno bodisi v pozitivno smer ipd. Kljub temu nas takšna porazdelitev prav nič ne moti, prav nasprotno. Ponovno se bomo vrnili k



Slika 5.5: Centroidi 50 gruĉ



Slika 5.6: Porazdelitev OHLC podatkov med gruĉe – 50 gruĉ



Slika 5.7: Porazdelitev OHLC podatkov med gruče – 20 gruč

naši analogiji, da so japonski svečniki besede – tako kot pri naravnem jeziku si tudi tu želimo imeti besede, ki so pogoste, kot tudi besede, ki so zelo redke. Tako kot tudi sami svoj vsakdan pogosto opišemo z besedno zvezo “nič posebnega”, zgolj v posebnih primerih pa uporabimo besedo “odlično”, ki nekako napoveduje bolj pestro nadaljevanje.

Iz podanih slik lahko sklepamo, da bo predlagan napovedni model deloval bolje pri večjem številu gruč japonskih svečnikov. Iz podatkov lahko enostavno razberemo, da je raznolikost japonskih svečnikov vsekakor večja v primeru 50 gruč kakor v primeru 20 gruč.

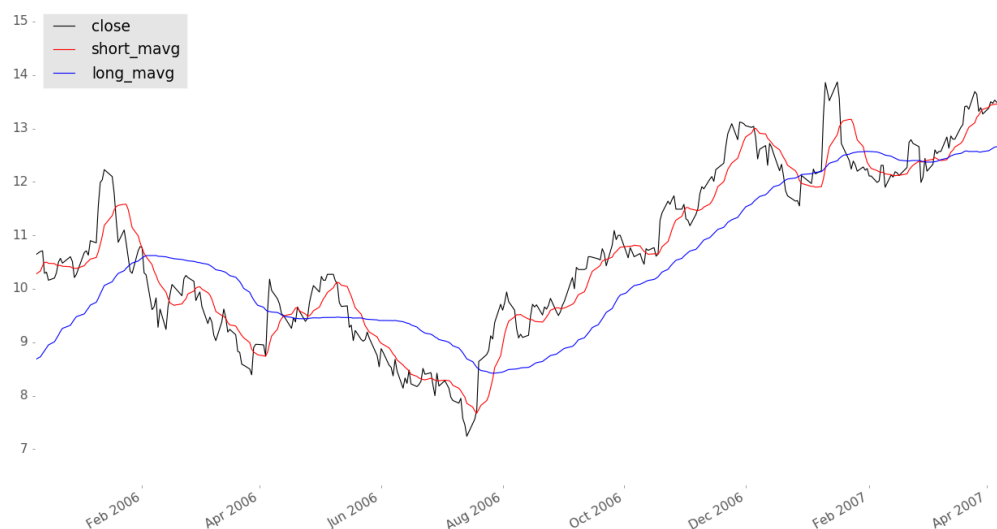
5.4 Testiranje napovednega modela

Do zdaj smo opisali tehnike in algoritme, ki smo jih uporabili. Podrobno smo si ogledali, kako smo zastavili napovedni model in trgovalno strategijo, ter testirali delovanje in kakovost le-te. Kot se izkaže, je predlagana trgovalna strategija kljub svoji enostavnosti dovolj robustna in se dobro obnese tudi pri ne ravno popolnih napovednih modelih – v določenih primerih tudi pri takšnih, kjer je verjetnost pravilne napovedi oziroma izbire trgovalne akcije med 50 – 60 %. V tem podpoglavju predstavimo rezultate testiranja napovednega modela, ki temelji na uporabi japonskih svečnikov in *Word2Vec* analize. Vsa testiranja smo opravili s pomočjo razvitega simulatorja in metrik uspešnosti, opisanih v poglavju 4. Najprej napovedujemo na podlagi preproste preslikave *OHLC* podatkov s pomočjo algoritma *Word2Vec*. V nadaljevanju podajamo rezultate za izboljšano različico napovednega modela, ki za posamezno napoved upošteva tudi zgodovinske podatke. Najprej pa si oglejmo še dva napovedna modela, katerih rezultati bodo služili za oceno uspešnosti predlaganega modela.

5.4.1 Tekoča povprečja – MA

Za potrebe primerjave rezultatov napovednega modela potrebujemo druge primerjalne modele. Eden izmed primerjalnih modelov bo preprost model *Kupi in zadrži*, drugi model, ki ga predstavimo v tem podpoglavju, pa imenujemo *Tekoča povprečja* (angl. *Moving Averages*) ali krajše MA.

Model, ki proži nakupne oziroma prodajne signale na podlagi tekočega povprečja, za svoje delovanje potrebuje dve tekoči povprečji – eno z dolgim ter drugo s kratkim oknom povprečenja. Dolžini obeh oken sta nastavljiva parametra modela. Osnovna logika delovanja je zelo preprosta: ko se vrednost kratkoročnega povprečja nahaja nad vrednostjo dolgoročnega povprečja, so delnice v fazi rasti, sicer so v fazi padca. Nakupne oziroma prodajne signale pa prožimo takrat, ko se vrednosti povprečij sekata. Prikaz gibanja tekočih povprečij v primerjavi z vrednostjo gibanja delnice prikazuje Slika 5.8.



Slika 5.8: Prikaz tekočih povprečij

Kako se tekoča povprečja obnesejo na podatkih delnic Coca-Cola, Microsoft in Apple v testnih in validacijskih obdobjih, prikazujeta Tabeli 5.8 in 5.9. Donosi so bili izračunani na podlagi \$1.000 začetne investicije in višini trgovalne provizije \$15.

| | Apple | Coca-Cola | Microsoft |
|-------------|-----------------|-----------------|-----------------|
| MA(10, 50) | \$5.474,59 (26) | - \$360,61 (25) | - \$183,20 (24) |
| MA(20, 50) | \$4.165,63 (22) | - \$446,87 (27) | - \$316,46 (24) |
| MA(45, 90) | \$3.121,95 (9) | - \$121,76 (12) | - \$538,27 (16) |
| MA(50, 100) | \$3.294,28 (7) | \$6,02 (12) | - \$539,60 (14) |
| MA(50, 200) | \$3.241,13 (5) | \$103,96 (7) | - \$278,36 (6) |

Tabela 5.8: Donosi in število transakcij za Tekoča povprečja v testnih obdobjih

Iz Tabele 5.8 razberemo, da predstavljeni model deluje v testnem obdobju slabše kakor model oz. strategija *Kupi in zadrži*, saj edini dobiček ustvari zgolj pri delnici podjetja Apple, kjer je ta za polovico manjši kakor pri stra-

| | Apple | Coca-Cola | Microsoft |
|-------------|-----------------|-----------------|-----------------|
| MA(10, 50) | \$1.142,18 (50) | - \$851,96 (63) | - \$201,09 (53) |
| MA(20, 50) | \$911,56 (42) | - \$710,03 (49) | - \$145,74 (41) |
| MA(45, 90) | \$2.819,21 (20) | \$181,45 (21) | - \$300,32 (24) |
| MA(50, 100) | \$2.856,26 (18) | \$69,39 (19) | - \$131,83 (22) |
| MA(50, 200) | \$6.007,51 (4) | \$275,06 (11) | \$517,25 (11) |

Tabela 5.9: Donosi in število transakcij za Tekoča povprečja v validacijskih obdobjih

tegiji *Kupi in zadrži*. V validacijskem obdobju pa zabeležimo velik uspeh pri kratkoročnem povprečju 50 dni ter dolgoročnem povprečju 100 ali 200 dni, kjer pri delnicah podjetja Apple ustvarimo tokrat kar 2-krat večji dobiček kot pri strategiji *Kupi in zadrži*, donosi pri Microsoftu in Coca-Coli pa so nekoliko nižji.

Na podlagi rezultatov primerjalnih modelov pridemo do zaključka, da bo s predlaganim modelom zelo težko doseči ali preseči donose najbolj enostavnega modela *Kupi in zadrži*.

5.4.2 MACD

Tretji napovedni model, katerega rezultate bomo primerjali s predlaganim napovednim modelom, je model MACD. Ta deluje podobno kakor model *Tekoča povprečja*, a s pomembno razliko. MACD temelji na uporabi uteženih tekočih povprečij – EMA (angl. *Exponential Moving Average*), ki so zelo podobna navadnim tekočim povprečjem, le da imajo zadnji dnevi večjo utež in posledično večji vpliv na rezultat povprečja.

MACD prav tako že definira kratka in dolga povprečja. Formula za izračun MACD je opisana v 5.1.

$$MACD = EMA(26) - EMA(12) \quad (5.1)$$

Signale prožimo tako, da izračunamo še 9-dnevni EMA linije MACD.

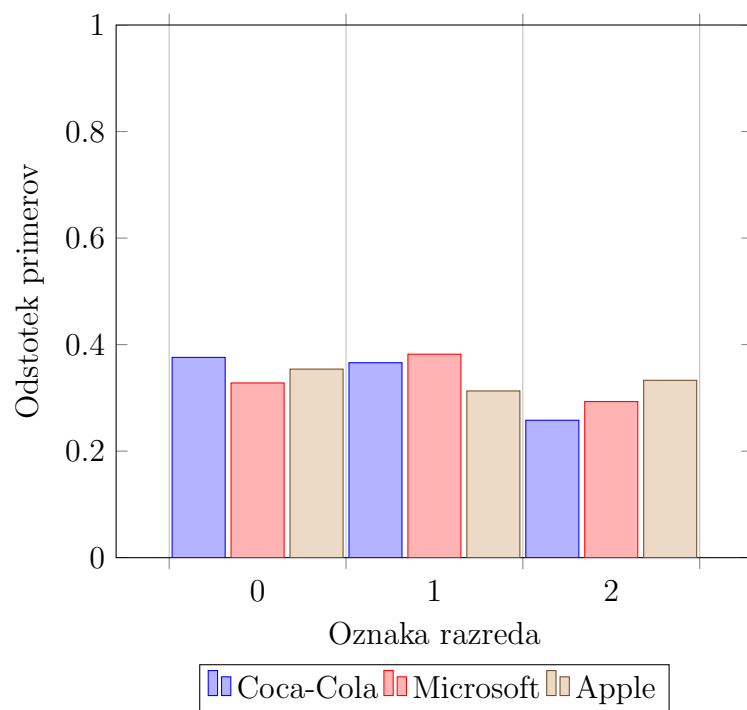
Podobno kakor pri modelu tekočih povprečij tudi pri MACD analiziramo prehode signalne linije čez MACD linijo. Dodaten signal, ki naznanjuje spremembo trenda, je tudi divergenca cene od MACD linije, katere v okviru tega dela nismo analizirali oziroma uporabili pri gradnji napovedi. MACD proži ogromno nakupnih in prodajnih signalov ter tako pri delnicah podjetja Microsoft in Coca-Cola zaključi z izgubo, saj so zneski plačanih provizij preprosto previsoki, da bi metoda delovala. Več uspeha pokaže šele pri višjih začetnih vložkih, katerih rezultate predstavimo na koncu tega poglavja.

5.4.3 Osnovno napovedovanje

Za potrebe analize si najprej oglejmo, kako se obnese čisto preprost napovedni model, kjer obstoječe *OHLC* podatke zgolj označimo z ustreznim razredom in uporabimo algoritem SoftMax. Primer naše matrike X in vektorja y je podan v Tabeli 5.10 – podatki *OHLC* so normirani po enačbi 3.3.

Preden se podamo v testiranje, si pogledjmo še porazdelitev učnih podatkov na Sliki 5.9, kjer smo razrede določili z optimalno vrednostjo parametra *period_frame_days* za posamezno delnico. Razberemo lahko, da so v učni množici vsi trije razredi zastopani približno enakomerno, kar je za naš klasifikacijski algoritem dobro, saj bo imel v fazi učenja dovolj primerov za vsakega izmed treh razredov. Na Sliki 5.10 za primerjavo podajamo še pravilno porazdelitev nad testnimi podatki, na Sliki 5.11 pa še na validacijski množici. Iz grafov razberemo, da sta porazdelitvi v testnem in validacijskem obdobju precej podobni, kjer prevladuje razred **2** – ni dovolj velike spremembe, ki bi naznanila akcijo.

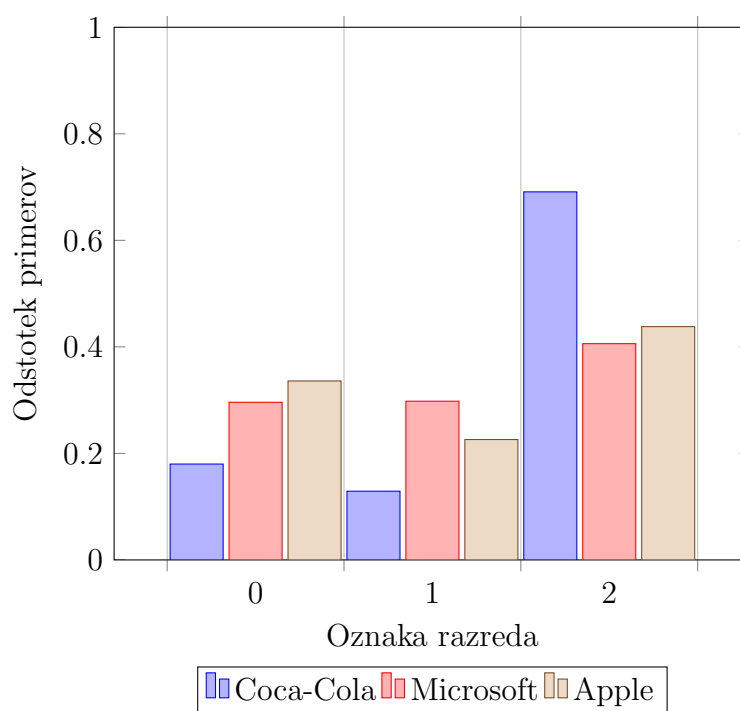
V tabelah 5.13, 5.12 in 5.11 podani rezultati prikazujejo, kako se osnovno napovedovanje obnese v praksi. Iz rezultatov v podanih tabelah lahko razberemo, da osnovno napovedovanje, brez obdelave z algoritmom *Word2Vec*, ne deluje prav dobro, kar smo tudi pričakovali. Klasifikacijske točnosti so izjemno nizke, s podrobnejšo analizo napovedi pa ugotovimo, da metoda v večini primerov vrača enak razred, bodisi razred 0 bodisi razred 1. Rezultati logaritmične izgube so prav tako zelo blizu primerjalne vrednosti 1,10. Iz ta-



Slika 5.9: Porazdelitev razredov v množici učnih podatkov

| O | H | L | C | y |
|--------|--------|--------|--------|---|
| 1,0000 | 1,0113 | 0,9978 | 1,0048 | 0 |
| 1,0000 | 1,0037 | 0,9773 | 0,9818 | 0 |
| 1,0000 | 1,0419 | 0,9938 | 1,0364 | 1 |
| 1,0000 | 1,0050 | 0,9557 | 0,9592 | 2 |
| 1,0000 | 1,0137 | 0,9975 | 1,0111 | 1 |

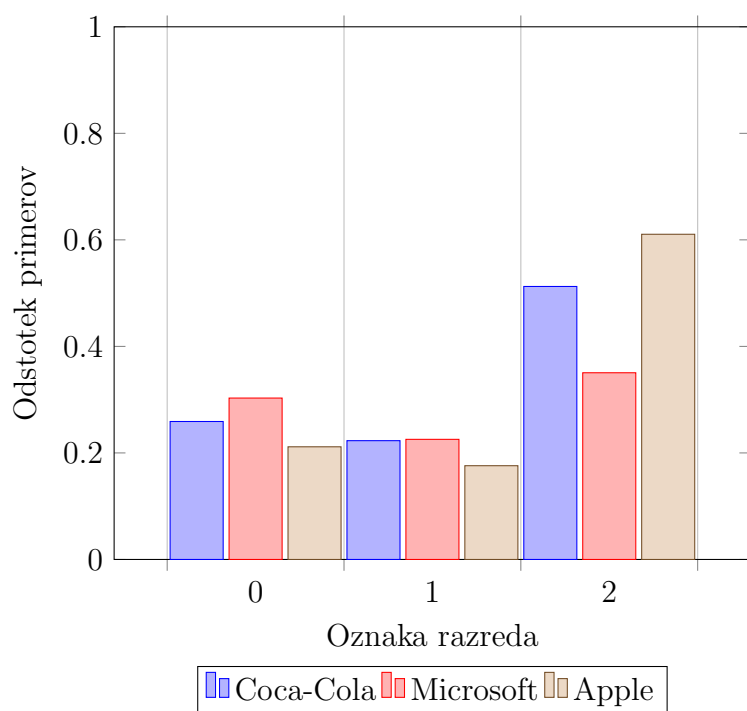
Tabela 5.10: Izsek matrike $X|y$



Slika 5.10: Porazdelitev razredov v množici testnih podatkov

| λ | logloss | klas. točnost | donos simulacije |
|--------------|---------|---------------|------------------|
| 1,0 | 1,18 | 18% | \$160,56 (1) |
| 0,1 | 1,23 | 18% | \$160,56 (1) |
| 0,01 | 1,24 | 18% | \$160,56 (1) |
| 0,001 | 1,24 | 18% | \$160,56 (1) |
| 0,0001 | 1,21 | 17% | - \$265,88 (23) |
| 0,00001 | 1,14 | 18% | - \$583,77 (43) |
| 0,000001 | 1,10 | 42% | - \$788,32 (51) |
| 0,0000001 | 1,09 | 45% | - \$915,76 (57) |
| 0,0000000001 | 1,09 | 45% | - \$898,59 (57) |

Tabela 5.11: Osnovno napovedovanje na delnici Coca-Cola v testnem obdobju



Slika 5.11: Porazdelitev razredov v množici validacijskih podatkov

| λ | logloss | klas. točnost | donos simulacije |
|--------------|---------|---------------|------------------|
| 1,0 | 1,11 | 29% | \$0,00 (0) |
| 0,1 | 1,12 | 29% | \$0,00 (0) |
| 0,01 | 1,12 | 29% | \$0,00 (0) |
| 0,001 | 1,11 | 29% | \$0,00 (0) |
| 0,0001 | 1,08 | 29% | \$151,09 (8) |
| 0,00001 | 1,03 | 48% | - \$66,84 (18) |
| 0,000001 | 1,02 | 49% | - \$15,30 (24) |
| 0,0000001 | 1,02 | 50% | - \$35,12 (28) |
| 0,0000000001 | 1,02 | 50% | - \$8,15 (26) |

Tabela 5.12: Osnovno napovedovanje na delnici Microsoft v testnem obdobju

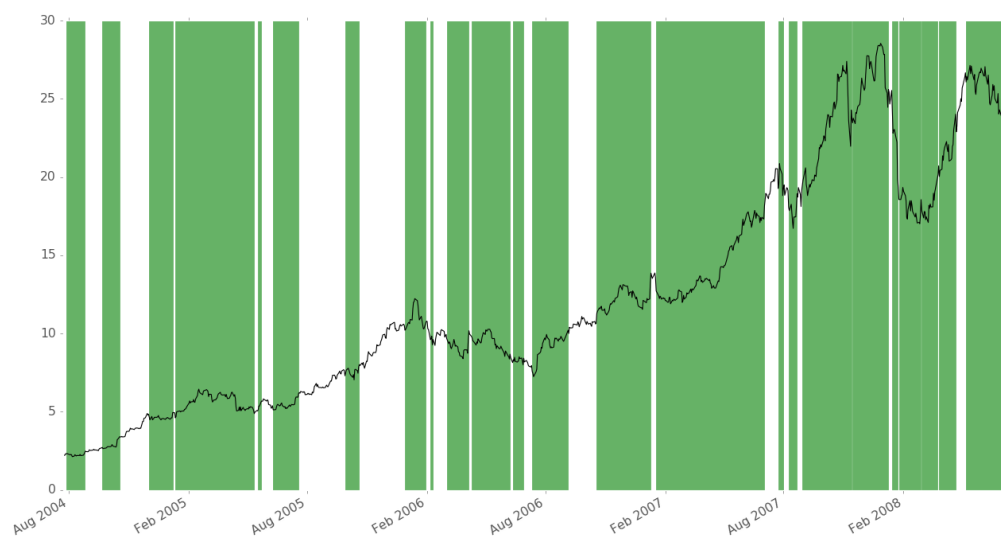
| λ | logloss | klas. točnost | donos simulacije |
|--------------|---------|---------------|------------------|
| 1,0 | 1,09 | 33% | \$9.872,22 (1) |
| 0,1 | 1,09 | 33% | \$9.872,22 (1) |
| 0,01 | 1,09 | 33% | \$9.872,22 (1) |
| 0,001 | 1,09 | 33% | \$9.872,22 (1) |
| 0,0001 | 1,07 | 45% | \$3.812,67 (49) |
| 0,00001 | 1,06 | 45% | \$1.979,62 (52) |
| 0,000001 | 1,06 | 45% | \$1.413,26 (52) |
| 0,0000001 | 1,06 | 45% | \$1.409,61 (52) |
| 0,0000000001 | 1,06 | 45% | \$1.409,61 (52) |

Tabela 5.13: Osnovno napovedovanje na delnici Apple v testnem obdobju

bel razberemo, da model za delnice Coca-Cole in Microsofta ustvari izgubo. Pozitiven rezultat simulacije pri podjetju Apple ne preseneti, saj je delnica v testnem obdobju, kot smo videli pri analizi podatkov, dosegla izjemno rast. Podrobnejši prikaz delovanja osnovnega napovedovanja, brez *Word2Vec* obdelave na delnici podjetja Apple, kjer je metoda dosegla pozitiven donos, prikazuje Slika 5.12. Če se vrnemo nazaj na Tabelo 5.1, kjer smo predstavili rezultate strategije *Kupi in zadrži*, vidimo, da je donos za delnico Apple tam mnogo večji, kar dodatno potrdi slabo delovanje osnovnega napovedovanja.

5.4.4 Napovedovanje z Word2Vec

V prvi iteraciji napovedovanja z razvitim napovednim modelom, ki temelji na *Word2Vec* obdelavi, smo podatke zgolj preslikali v njihove vektorske predstavitve in se podali v napovedovanje. V postopku napovedovanja, po zaznavanju japonskih svečnikov in tvorbi stavkov, smo le-te posredovali algoritmu *Word2vec*, ki se je potem naučil kontekst besed japonskih svečnikov. Posamezen *OHLC* podatek je bil tako preslikan v zgolj enega od japonskih svečnikov, te pa smo nato s pomočjo *Word2Vec* razširili v vektorje dimenzije 100. Izsek tako pridobljene matrike podajamo v Tabeli 5.14, matriko upo-



Slika 5.12: Delovanje osnovnega napovednega modela na delnicah podjetja Apple

| X | | | | | | | | y |
|-------------|-------------|-------------|-------------|------------|-------------|-----|-------------|---|
| -0,93611144 | 0,61355882 | 0,1940219 | 0,3965361 | 0,6306261 | 0,8294799 | ... | 0,44962329 | 0 |
| -1,31536217 | 0,28227576 | -0,03256915 | -0,95867816 | 1,45770603 | -1,01643918 | ... | -0,06653133 | 0 |
| 1,28682947 | 0,5157703 | -0,80432536 | -0,6182747 | 0,06461049 | 0,21013272 | ... | 1,395083 | 1 |
| 0,19670276 | -0,47933341 | 1,57494792 | -0,08408138 | 0,09630984 | 0,2977032 | ... | -0,04894764 | 2 |

Tabela 5.14: Izsek matrike $X|y$ po obdelavi z Word2Vec

rabimo za osnovo klasifikacijske metode SoftMax. Naši primeri imajo sedaj 100 dimenzij, napovedujemo pa njihovo pripadnost enemu izmed 3 razredov ter testiramo samo delovanje napovednega modela.

Tehten razmislek nam pokaže, da bo naš model – ne glede na izbiro parametrov – deloval slabše, kakor deluje osnovno napovedovanje brez uporabe *Word2Vec* preslikave. Da je res tako, lahko razberemo iz samega opisa postopka – podatke *OHLC* po obdelavi zgolj preslikamo v enega izmed n japonskih svečnikov. Če so bili naši učni primeri v matriki X med seboj praktično vsi različni, imamo sedaj zgolj n različnih vektorjev dimenzije 100. Poleg tega imamo seveda za enake primere x v učni množici različne pravilne vrednosti v pripadajočem vektorju y . Da takšen način napovedovanja ne bo

| λ | logloss | klas. točnost | donos simulacije |
|--------------|---------|---------------|------------------|
| 1,0 | 1,10 | 33% | \$9.872,22 (1) |
| 0,1 | 1,09 | 33% | \$9.872,22 (1) |
| 0,01 | 1,08 | 45% | \$9.822,99 (1) |
| 0,001 | 1,06 | 45% | \$7.171,31 (9) |
| 0,0001 | 1,06 | 45% | \$5.411,06 (33) |
| 0,00001 | 1,06 | 42% | - \$825,39 (66) |
| 0,000001 | 1,06 | 42% | - \$57,16 (26) |
| 0,0000001 | 1,06 | 42% | - \$57,16 (26) |
| 0,0000000001 | 1,06 | 42% | - \$57,16 (26) |

Tabela 5.15: Napovedovanje W2V brez združevanja dni na podatkih delnice Apple

deloval, je torej jasno – vseeno pa prilagamo Tabelo 5.15, kjer lahko rezultate primerjamo s Tabelo 5.13. Vidimo, da so rezultati v primerjavi z navadnim napovedovanjem slabši, kar je v skladu z našim sklepanjem. Parametri metode sicer zaradi zgoraj opisanega niti niso tako zelo pomembni, a vseeno velja omeniti, da smo število gruč in dolžino stavkov nastavili na 20.

5.4.5 Združevanje s seštevanjem

Spoznanje, da osnovno napovedovanje brez kakršnihkoli preslikav deluje slabo, ni bilo nikakršno presenečenje. Spoznanje, da zastavljeni model v sami osnovi ne bo prinesel nikakršne izboljšave, morda ni ravno preveč obetavno, a kljub temu smo identificirali vzrok slabega delovanja.

Cilj modela je obravnava podatkov *OHLC* kot besed z algoritmom *Word2Vec*, podobno kot so to storili Mikolov in sodelavci [4, 15], a ne na področju trgovanja z delnicami. V poglavju z opisom algoritma *Word2Vec* smo pokazali, da se po obdelavi sistem nauči konteksta posameznih besed in je tako sposoben naprednih sklepanj, kot prikazuje primer 5.2.

| λ | merge_days | n_clusters | sentence_length | logloss | klas. točnost | Donos simulacije |
|----------------|------------|------------|-----------------|-------------|---------------|-------------------------|
| * 0,001 | 2 | 50 | 30 | 1,05 | 46% | \$16.732,49 (23) |
| 0,001 | 3 | 70 | 30 | 1,05 | 46% | \$14.949,12 (15) |
| 0,001 | 4 | 40 | 30 | 1,05 | 45% | \$14.074,28 (17) |
| 0,0001 | 1 | 50 | 5 | 1,07 | 44% | \$14.785,35 (17) |
| 0,0001 | 3 | 50 | 20 | 1,05 | 46% | \$14.530,55 (22) |
| 0,0001 | 4 | 50 | 20 | 1,06 | 45% | \$14.689,72 (18) |
| 0,0001 | 6 | 50 | 5 | 1,06 | 46% | \$16.295,65 (22) |
| 0,0001 | 6 | 70 | 10 | 1,06 | 44% | \$14.979,03 (21) |
| 0,0001 | 7 | 50 | 10 | 1,06 | 46% | \$16.457,44 (18) |
| 0,0001 | 7 | 70 | 10 | 1,06 | 44% | \$14.412,62 (19) |
| 0,0001 | 8 | 50 | 5 | 1,06 | 45% | \$15.398,47 (22) |
| 0,0001 | 9 | 50 | 5 | 1,06 | 44% | \$14.379,37 (20) |
| 0,0001 | 9 | 70 | 10 | 1,07 | 43% | \$14.877,31 (15) |
| 0,000001 | 4 | 70 | 20 | 1,10 | 41% | \$14.671,77 (52) |

Tabela 5.16: Napovedovanje W2V z združevanjem na podatkih delnice podjetja Apple

$$\text{Kralj} - \text{Moški} + \text{Ženska} = \text{Kraljica} \quad (5.2)$$

Podatke *OHLC* pretvorimo, kot prej, v vektorje ter nato vektorsko seštejemo zadnjih n dni. Rezultat vektorskega seštevanja vstavimo kot primer v matriko X , ki je po obliki podobna matriki v Tabeli 5.14.

V Tabeli 5.16 podajamo nekaj izmed bolj zanimivih rezultatov za delnice podjetja Apple, kjer smo modelu podali sledeče parametre:

- λ – stopnja regularizacije
- *merge_days* – koliko zaporednih dni iz preteklosti vektorsko seštejemo
- *n_clusters* – koliko različnih japonskih svečnikov želimo imeti v fazi gručenja
- *sentence_length* – dolžina tvorjenih stavkov

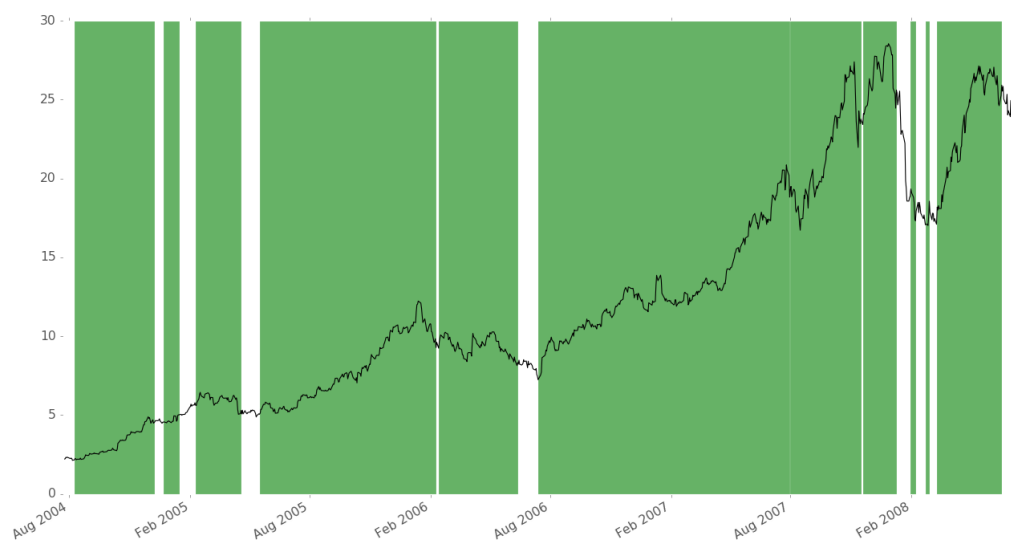
Vidimo lahko, da so donosi za izbrane parametre kar precej višji, kakor pri strategiji *Kupi in zadrži*, saj predlagani napovedni model ustvari kar \$16.732,49 dobička iz začetnih \$1.000 pri 23 proženih akcijah. Klasifikacijska točnost in logaritmična izguba sta precej slabi, pravilno napovemo trend za

zgolj okoli 45% trgovalnih dni, verjetnosti pripadnosti razredom pa so zelo blizu primerjalne vrednosti 1,10. Seveda nas izmed vseh treh mer najbolj zanima višina donosa, saj je – kot smo spoznali že prej – logaritmična izguba lahko zelo visoka, če za zgolj en primer z veliko verjetnostjo identificiramo napačen trend.

Ker vemo, da delnice podjetja Apple v splošnem znotraj testnega obdobja rastejo, si zato podrobneje pogledjmo še graf lastništva delnic, kjer lahko podrobneje analiziramo delovanje modela na Sliki 5.13. Kot lahko na hitro razberemo, je v našem primeru model zaznal obdobje največjega padca proti koncu grafa. Vsekakor pa ni popoln, saj vidimo, da je kar veliko padcev povsem spregledal. Zato podajamo še graf, kjer je model sicer dosegel nekoliko slabši splošni donos \$14.671,77, a kot lahko razberemo iz grafa na Sliki 5.14, veliko bolje zaznal negativne trende. Graf je bil narisani pri izbiri parametrov, kjer je model opravil največ trgovalnih akcij:

1. *merge_days*: 4
2. *n_clusters*: 70
3. *sentence_length*: 20
4. λ : 0,000001

Rezultati na delnicah podjetja Apple kažejo za predlagani napovedni model velik potencial, pogledjmo si še rezultate za delnice podjetja Microsoft v tabeli 5.17 in podjetja Coca-Cola v tabeli 5.18. Pri delnicah podjetja Microsoft v testnem obdobju smo videli, da je smer gibanja delnice v negativnem trendu, kar se kaže tudi v rezultatu strategije *Kupi in zadrži*, ki zaključi trgovanje z \$292,70 izgube. Iz tabel lahko razberemo, da predlagani napovedni model zaključi enako obdobje z 200 do 300 dolarji dobička, kjer smo za provizije zapravili vsega skupaj med 150 in 630 dolarji. Pogledjmo si še grafično vizualizacijo delovanja modela nad delnicami podjetja Microsoft na Sliki 5.15, kjer prikažemo delovanje pri največjem donosu. Razberemo lahko,

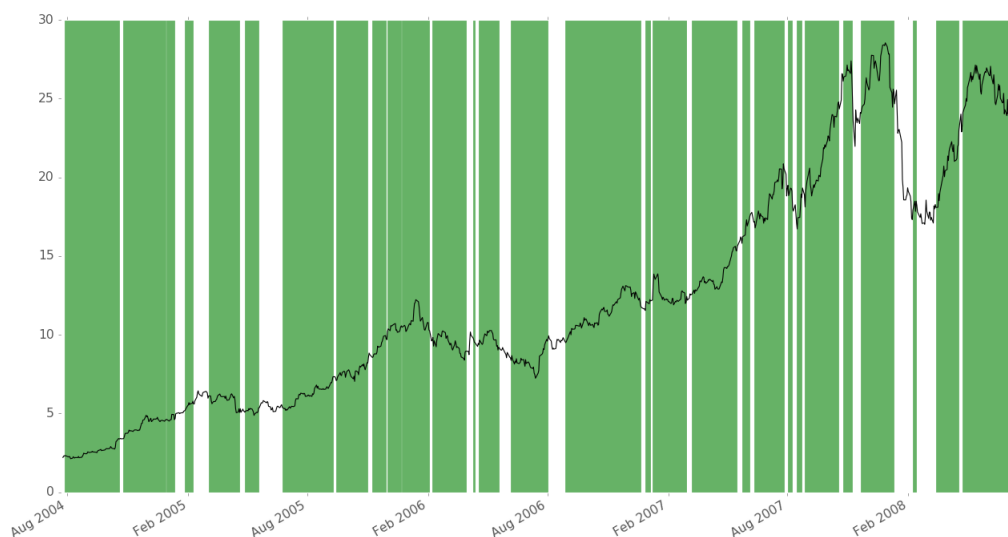


Slika 5.13: Delovanje W2V napovednega modela na delnicah podjetja Apple v testnem obdobju

da je tudi tokrat naš model precej dobro zaznal splošen trend gibanja. Dodatno lahko vidimo, da so rezultati metrik logaritmične izgube pri delnicah podjetja Microsoft mnogo boljši od primerjalne vrednosti 1,10. Tudi sam odstotek pravilnih napovedi se povzpne nad 50 % (glej tabelo 5.17).

Pri pregledu grafa na Sliki 5.16, ki prikazuje delovanje na delnici Coca-Cole, vidimo, da model sicer zazna začetni negativni trend, kasneje pa ne proži veliko akcij – vsega skupaj zgolj 13 akcij, vendar še vedno konča z rahlo prednostjo pred strategijo *Kupi in zadrži*. Metrike logaritmične izgube in klasifikacijske točnosti pri delnicah Coca-Cole dosegajo zelo dobre rezultate, pravilno klasificiramo preko 60 % primerov, logaritmična izguba pa je v vseh primerih nižja od 1, kar je veliko bolje od primerjalne vrednosti 1,10.

Če primerjamo rezultate v vseh treh tabelah, vidimo, da naš model deluje zelo dobro v smislu donosa. Če danim donosom prištejemo še zneske plačanih provizij, vidimo, da so rezultati v praksi zelo dobri. Ob analizi vrednosti parametrov ugotovimo, da smo pravilno sklepali, da bo model deloval bolje pri višjem številu gruč japonskih svečnikov – pri vseh treh delnicah se je



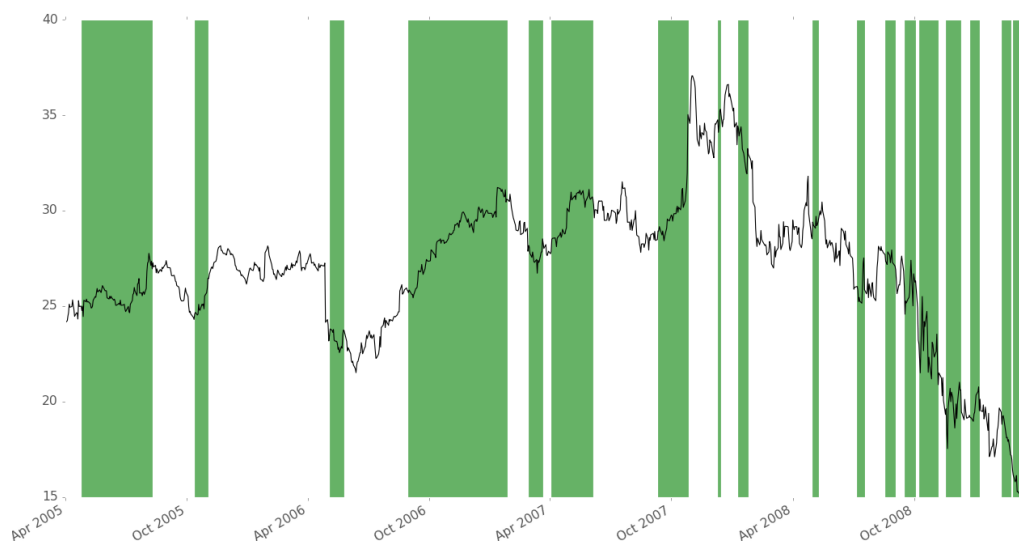
Slika 5.14: Delovanje W2V napovednega modela na delnicah podjetja Apple v testnem obdobju

| λ | merge_days | n_clusters | sentence_length | logloss | klas. točnost | Donos simulacije |
|---------------|------------|------------|-----------------|-------------|---------------|----------------------|
| 0,001 | 9 | 60 | 5 | 1,02 | 50% | \$252,53 (12) |
| 0,001 | 10 | 60 | 5 | 1,01 | 51% | \$252,79 (10) |
| 0,001 | 30 | 90 | 30 | 0,98 | 52% | \$231,28 (11) |
| 0,0001 | 8 | 40 | 10 | 1,01 | 51% | \$269,10 (42) |
| 0,0001 | 9 | 40 | 25 | 1,00 | 52% | \$273,67 (42) |
| 0,0001 | 9 | 40 | 20 | 1,00 | 51% | \$357,37 (36) |
| 0,0001 | 9 | 40 | 25 | 1,00 | 52% | \$273,67 (42) |
| * 0,0001 | 10 | 40 | 20 | 1,00 | 53% | \$340,58 (42) |

Tabela 5.17: Napovedovanje W2V z združevanjem za delnice podjetja Microsoft

| λ | merge_days | n_clusters | sentence_length | logloss | klas. točnost | Donos simulacije |
|--------------|------------|------------|-----------------|-------------|---------------|----------------------|
| 0,001 | 6 | 90 | 20 | 0,97 | 57% | \$148,54 (11) |
| 0,001 | 7 | 90 | 20 | 0,96 | 60% | \$143,04 (13) |
| 0,001 | 8 | 90 | 20 | 0,96 | 61% | \$200,11 (11) |
| 0,001 | 20 | 20 | 25 | 0,97 | 59% | \$158,47 (11) |
| 0,001 | 30 | 60 | 5 | 0,96 | 64% | \$242,70 (13) |
| * 0,001 | 30 | 60 | 10 | 0,94 | 63% | \$230,60 (13) |

Tabela 5.18: Napovedovanje W2V z združevanjem za delnice podjetja Coca-Cola



Slika 5.15: Delovanje W2V napovednega modela na delnicah podjetja Microsoft v testnem obdobju

model namreč najbolje odrezal s 40 ali več gručami. Zanimivi so tudi ostali parametri, saj pri delnicah Apple in Microsoft vidimo, da je vrednost parametra *merge_days* zelo podobna vrednosti *period_frame_days*, medtem ko se ta vrednost pri delnici podjetja Coca-Cola (pri najboljšem modelu) močno razlikuje.

| Ime | Kupi in zadrži | MA(50,100) | MACD | W2V |
|-----------|----------------|-----------------|-----------------|-----------------|
| Apple | \$2.927,13 | \$2.856,26 (18) | - \$940,73 (72) | \$3.508,22 (22) |
| Microsoft | \$1.938,35 | - \$131,83 (22) | - \$979,40 (82) | - \$303,17 (32) |
| Coca-Cola | \$519,59 | \$69,39 (19) | - \$965,93 (52) | \$468,00 (15) |
| Povprečno | \$1.795,02 | \$913,27 (20) | - \$961,92 (86) | \$1.224,35 (23) |

Tabela 5.19: Donosi napovednih modelov za podane delnice v validacijskem obdobju pri začetni investiciji \$1.000

V Tabeli 5.19 podajamo še rezultate v validacijskem obdobju s parametri, ki so v testnem obdobju prinesli najvišje donose. Razberemo lahko, da v



Slika 5.16: Delovanje W2V napovednega modela na delnicah podjetja Coca-Cola v testnem obdobju

primeru delnic Apple in Coca-Cola tudi v validacijskem obdobju pridelamo dobiček. Za delnice podjetja Microsoft pa, kljub splošni rasti v validacijskem obdobju, model, ki se je najbolj odrezal na testnem obdobju, tu ustvari izgubo \$303,17. Seveda moramo ponovno upoštevati, da je napovedni model za vsako proženo akcijo (število akcij je navedeno v oklepaju) plačal \$15 provizije. Grafe, kako se napovedni modeli odrežejo v validacijskem obdobju, podajamo na Slikah 5.17, 5.18 in 5.19. Kot lahko razberemo iz grafov, vidimo, da se napovedni modeli zelo podobno obnašajo, dajejo podobne napovedi pri vseh treh delnicah kakor v testnem obdobju. Tako lahko vidimo na delnicah podjetja Microsoft, ki je prav tako kot podjetje Apple, v validacijskem obdobju v obdobju rasti, da je napovedni model precej bolj previden, kar se kaže tudi v slabšem donosu.

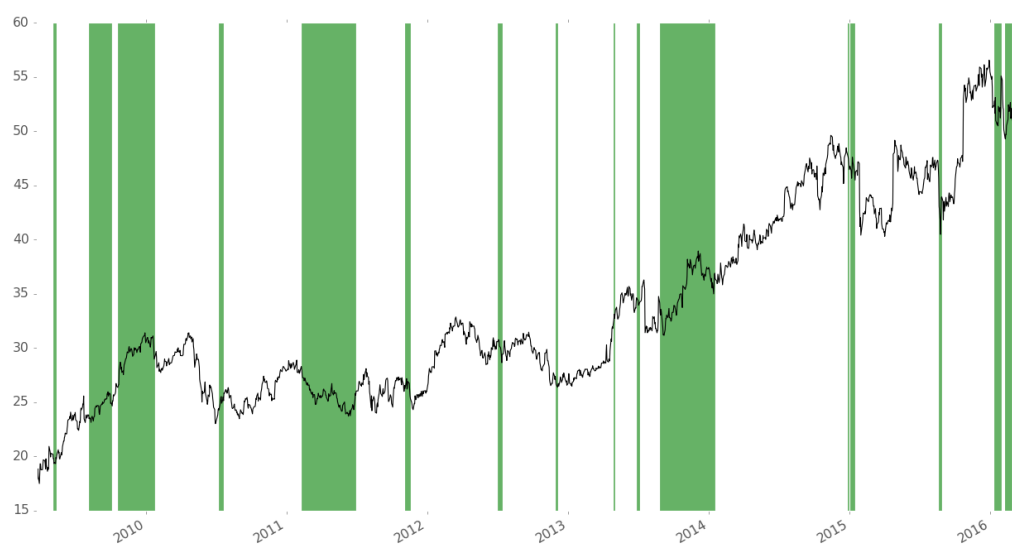


Slika 5.17: Delovanje W2V modela v validacijskem obdobju za podatke delnice Apple

Poglejmo si za primerjavo še, kako so se obnesli modeli, ki so v testnem obdobju pokazali najmanjšo logaritmčno izgubo in najvišjo klasifikacijsko točnost – Tabela 5.20. Modeli, katerih parametri so bili izbrani, so v tabelah 5.16, 5.17 in 5.18 označeni z (*). Iz tabele lahko tokrat razberemo, da se predlagani napovedni model za vse tri delnice zelo dobro obnese tudi v validacijskem obdobju.

| Ime | Kupi in zadrži | MA(50,100) | MACD | W2V |
|-----------|----------------|-----------------|-----------------|-----------------|
| Apple | \$2.927,13 | \$2.856,26 (18) | - \$940,73 (72) | \$3.508,22 (22) |
| Microsoft | \$1.938,35 | - \$131,83 (22) | - \$979,40 (82) | \$585,65 (20) |
| Coca-Cola | \$519,59 | \$69,39 (19) | - \$965,93 (52) | \$798,65 (9) |
| Povprečno | \$1.795,02 | \$913,27 (20) | - \$961,92 (86) | \$1.630,84 (17) |

Tabela 5.20: Donosi napovednih modelov za podane delnice v validacijskem obdobju pri začetni investiciji \$1.000



Slika 5.18: Delovanje W2V modela v validacijskem obdobju za podatke delnice Microsoft



Slika 5.19: Delovanje W2V modela v validacijskem obdobju za podatke delnice Coca-Cola

5.4.6 Primerjava metod pri višjih zneskih začetnega kapitala

Za zaključek si pogledajmo še, kako se primerjalni modeli *Kupi in zadrži*, *Tekoča povprečja* in *MACD* ter predlagani napovedni model obnesejo v testnih in validacijskih množicah pri znesku začetne investicije \$10.000, kjer je znesek provizije zanemarljivo majhen ter tako ne vpliva veliko na končni rezultat simulacije. Ostale metrike predlaganega napovednega modela so enake kakor v tabelah 5.16, 5.17 in 5.18. Rezultati simulacije za omenjene napovedne modele pri višji začetni investiciji na testnih podatkih so podani v Tabeli 5.21.

| Ime | Kupi in zadrži | MA(50,100) | MACD | W2V |
|-----------|----------------|-------------------|-------------------|-------------------|
| Apple | \$102.557,08 | \$34.915,34 (7) | \$46.452,72 (67) | \$182.938,35 (23) |
| Microsoft | - \$2.927,03 | - \$4.140,42 (14) | - \$3.261,15 (73) | \$11.109,06 (42) |
| Coca-Cola | \$1.996,82 | \$2.066,74 (12) | - \$1.108,05 (86) | \$4.360,76 (13) |
| Povprečno | \$33.875,62 | \$10.947,21 (11) | \$14.027,84 (75) | \$66.136,05 (26) |

Tabela 5.21: Donosi napovednih modelov za podane delnice v testnem obdobju pri začetni investiciji \$10.000

Podrobnejše rezultate napovednih modelov na validacijski množici podamo v Tabeli 5.22. Iz tabel lahko zaključimo, da je predlagani napovedni model v testni množici boljši od vseh treh primerjalnih modelov. Pri delnici podjetja Microsoft, ki v testnem obdobju izgublja na vrednosti, pa je predlagani model pravzaprav edini, ki ustvari dobiček. Predlagani model se v validacijski množici obnese nekoliko slabše, saj pri delnicah podjetij Coca-Cola in Microsoft ustvari manjši dobiček kakor primerjalni model *Kupi in zadrži*, vendar v povprečju še vedno večji skupni donos kakor vsi ostali napovedni modeli.

| Ime | Kupi in zadrži | MA(50,100) | MACD | W2V |
|-----------|----------------|------------------|--------------------|------------------|
| Apple | \$28.611,11 | \$32.339,63 (18) | \$6.619,31 (158) | \$57.543,47 (22) |
| Microsoft | \$20.316,42 | \$1.809,31 (22) | \$2.477,12 (133) | \$10.603,90 (9) |
| Coca-Cola | \$5.547,81 | \$3.583,26 (19) | - \$4.220,57 (183) | \$3.163,32 (36) |
| Povprečno | \$18.158,45 | \$12.577,40 (20) | \$1.625,28 (158) | \$23.770,19 (22) |

Tabela 5.22: Donosi napovednih modelov za podane delnice v validacijskem obdobju pri začetni investiciji \$10.000

5.5 Statistična primerjava

Analiza rezultatov na delnicah podjetij Coca-Cola, Microsoft in Apple je pokazala, da predlagani napovedni model v sodelovanju s predlagano trgovalno strategijo deluje bolje od vseh primerjalnih modelov v testnem obdobju. Nekoliko slabše rezultate smo dosegli v validacijskem obdobju, a je bil povprečen donos predlaganega modela še vedno višji od primerjanih modelov. Ker je vzorec treh delnic absolutno premajhen za konkretne zaključke o uspešnosti delovanja, v nadaljevanju opravimo obsežnejše testiranje na večjem naboru podatkov ter opravimo statistično primerjavo s pomočjo *Wilcoxonovega testa s predznačenimi rangi*.

5.5.1 Russell Top 50

Postopek napovedovanja smo opravili za vsako delnico posebej. Tako smo za vsako izmed 50 delnic napovedni model učili na učni množici trenutne delnice ter iskali nabor parametrov, ki prinašajo najvišji donos v testnem obdobju. Ko smo opravili učenje za vsako izmed 50 delnic, smo model z ustreznimi parametri pognali še na validacijski množici. V Tabeli 5.23 so podani povprečni donosi predlaganega modela kot tudi primerjalnih modelov oz. strategij *Kupi in zadrži*, *Tekoča povprečja* in *MACD* za testno in validacijsko obdobje. Iz tabele lahko razberemo, da je povprečen donos predlaganega modela na celotnem indeksu v testnem obdobju deloval mnogo bolje od preostalih, medtem ko v validacijskem obdobju pridelamo višji povprečni donos s strategijo *Kupi in zadrži*. Podrobnejši rezultati donosov za posamezno delnico so podani v tabelah 5.24 in 5.25. Kot lahko iz tabel razberemo predlagani napovedni model v testnem obdobju z ustreznimi parametri pri prav vseh delnicah ustvari dobiček. V validacijskem obdobju napovedni model pri nekaterih delnicah ustvari tudi izgubo a je takšnih delnic zelo malo.

| Ime | Kupi in zadrži | MA(50,100) | MACD | W2V |
|-------------------|----------------|------------|------------|-------------|
| Testno obdobje | \$2.818,98 | \$1.073,06 | - \$482,04 | \$11.725,25 |
| Validacijsko obd. | \$16.590,83 | \$6.238,43 | \$395,10 | \$10.324,24 |

Tabela 5.23: Povprečni donosi napovednih modelov na delnicah Russel Top 50 pri začetni investiciji \$10.000

5.5.2 Wilcoxonov test s predznačenimi rangi

Ker nam povprečje ne pove, ali predlagani model tudi statistično značilno deluje bolje od primerjalnih modelov oz. strategij, želimo opraviti *Wilcoxonov test s predznačenimi rangi* ter tako potrditi ali zavrniti našo hipotezo, da predlagani napovedni model deluje statistično značilno bolje.

Postopek preverjanja statistične značilnosti je sledeč:

1. izračunamo razlike donosov d primerjanih modelov za posamezno delnico,
2. izračunamo absolutne vrednosti d ter jih razvrstimo od najmanjšega do največjega,
3. rangiramo vrednosti d ter jim pripišemo predznak vrednosti d ,
4. seštejemo pozitivne range v vrednost \mathbf{W}^+ ter negativne range v vrednost \mathbf{W}^- ,
5. absolutna vrednost manjše vsote je statistika metode \mathbf{W} ,
6. preverimo statistično značilnost s pomočjo približka **z-vrednosti** po enačbi 5.3 ter poiščemo pripadajočo **p-vrednost**.

$$z \approx \frac{W - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}} \quad (5.3)$$

Standardna kritična meja α za sprejem hipoteze, da predlagani model deluje bolje, je 0,05. To pomeni, da bo predlagani model statistično značilno deloval bolje v primeru, da bi z naključnimi podatki pridobili boljše rezultate od primerjalnega modela v manj kot 5 %. Našo hipotezo sprejmemo za **p-vrednosti**, ki so manjše od 0,05.

5.5.3 Rezultati

Izračunane **p-vrednosti** in statistike **W** so podane v Tabeli 5.26. Rezultati kažejo, da se predlagani model s pravilno izbiro parametrov izkaže statistično značilno bolje od vseh ostalih modelov z veliko stopnjo zaupanja, saj so **p-vrednosti** zelo majhne. Takšne rezultate smo uspeli pridobiti znotraj testnega obdobja. V validacijskem obdobju pridemo do zanimivih rezultatov. Predlagani model statistično značilno deluje bolje zgolj v primerjavi z *MACD* ter *Tekoča povprečja – MA(50, 100)*, kjer pridobimo **p-vrednost** 0,021, kar je za več kot polovico manj od mejne vrednosti α 0,05. Pri primerjavi s *Kupi in zadrži* smo že z enostavnim povprečjem videli, da predlagani model prinaša slabše donose. Da *Kupi in zadrži* v validacijskem obdobju deluje tudi statistično boljše od predlaganega modela smo potrdili z dodatnim testom ter pridobili p-vrednost manjšo od 0,0001.

Zaključimo lahko, da predlagani napovedni model z dobro izbiro parametrov statistično značilno deluje bolje od primerjanih modelov oz. strategij, a parametri, ki dobro delujejo v testnem obdobju, ne delujejo enako dobro tudi v kasnejšem validacijskem obdobju.

| Delnica | Kupi in zadrži | MA(50,100) | MACD | W2V |
|---------|----------------|--------------|--------------|-------------|
| VZ | - \$3.039,61 | \$647,66 | - \$2.990,29 | \$3.080,84 |
| T | - \$359,55 | - \$42,32 | - \$1.280,35 | \$8.273,56 |
| UNH | - \$3.769,27 | - \$288,51 | - \$3.398,75 | \$5.050,80 |
| AMGN | - \$291,37 | \$252,85 | \$3.227,64 | \$7.668,43 |
| GE | - \$3.927,61 | - \$2.472,75 | - \$3.128,26 | \$2.956,75 |
| CELG | \$25.708,81 | \$5.608,72 | \$12.781,08 | \$50.882,06 |
| CMCSA | - \$542,23 | \$982,66 | - \$2.407,12 | \$2.889,15 |
| KO | \$2.519,05 | \$3.778,97 | - \$1.866,26 | \$5.813,29 |
| MCD | \$8.955,17 | \$1.114,14 | \$898,75 | \$12.868,28 |
| AGN | \$3.441,65 | \$2.536,50 | \$747,18 | \$5.693,08 |
| QCOM | - \$772,94 | \$1.228,42 | - \$3.761,31 | \$9.237,34 |
| SLB | \$11.208,42 | \$8.533,32 | \$596,46 | \$3.3575,8 |
| HD | - \$4.491,7 | - \$5.002,76 | - \$4.843,12 | \$2.520,49 |
| BAC | - \$4.641,57 | - \$4.017,49 | - \$2.588,89 | \$5.880,18 |
| PFE | - \$3.913,79 | - \$3.236,17 | - \$4.358,06 | \$3.252,18 |
| WFC | \$309,97 | - \$3.488,26 | - \$6.321,32 | \$6.627,38 |
| CVX | \$1.686,49 | - \$903,09 | \$5.414,68 | \$9.172,12 |
| UTX | \$1.132,83 | \$301,21 | - \$1.690,97 | \$5.169,85 |
| MDT | \$1.543,71 | \$81,96 | - \$3.058,66 | \$5.608,06 |
| HON | - \$138,26 | - \$366,97 | \$147,02 | \$6.283,95 |
| BMJ | - \$1.934,93 | - \$1.997,37 | - \$2.820,40 | \$5.111,98 |
| BA | \$14,22 | \$2.171,52 | - \$1.990,97 | \$6.615,65 |
| IBM | \$702,70 | - \$593,99 | - \$367,11 | \$5.488,49 |
| WMT | \$410,02 | - \$2.501,45 | - \$1.724,35 | \$6.359,24 |
| AAPL | \$31.114,37 | \$17.420,48 | \$26.956,85 | \$97.776,72 |
| MSFT | - \$1.231,96 | - \$4.080,51 | - \$674,72 | \$8.067,93 |
| BRKB | \$4.642,86 | \$2.359,40 | - \$2.075,03 | \$8.411,10 |
| MA | \$12.182,20 | \$9.519,22 | \$2.831,36 | \$21.774,12 |
| DIS | \$674,97 | \$2.110,20 | - \$818,91 | \$1.479,87 |
| V | \$4.439,16 | \$4.203,69 | - \$2.030,36 | \$8.327,75 |
| MMM | - \$2.045,36 | - \$3.388,90 | - \$2.325,36 | \$2.073,84 |
| PM | \$8.102,11 | \$3.265,02 | \$2.208,59 | \$13.333,25 |
| INTC | - \$1.856,66 | - \$1.277,88 | - \$830,56 | \$4.716,03 |
| CSCO | \$74,87 | - \$1.609,55 | - \$3.946,78 | \$8.631,76 |
| PG | \$2.500,00 | - \$991,19 | - \$934,31 | \$8.266,96 |
| GOOGL | - \$1.031,18 | - \$230,00 | \$2.288,33 | \$10.551,24 |
| UNP | \$9.815,55 | \$1.029,44 | \$2.830,19 | \$18.177,28 |
| JNJ | \$1.038,69 | - \$641,97 | \$20,70 | \$0,00 |
| MRK | - \$278,32 | \$759,97 | - \$2.167,89 | \$13.128,04 |
| XOM | \$5.746,22 | \$2.277,44 | - \$771,99 | \$6.934,31 |
| MO | - \$6.182,24 | \$616,95 | - \$8.350,84 | \$2.346,59 |
| AMZN | \$6.610,64 | \$5.827,54 | \$6.074,77 | \$35.389,87 |
| ABBV | \$2.900,97 | \$3.322,96 | \$1.512,46 | \$5.227,56 |
| GILD | \$11.832,06 | \$8.258,70 | - \$5.391,78 | \$22.280,68 |
| ORCL | \$3.631,19 | \$1.747,77 | - \$1.572,96 | \$8.642,77 |
| FB | \$18.046,26 | \$4.560,36 | \$2.071,40 | \$23.657,48 |
| C | - \$6.524,43 | - \$2.595 | - \$5.441,80 | \$8.239,60 |
| CVS | \$3.344,02 | \$833,63 | - \$3.662,15 | \$13.220,15 |
| PEP | \$3.235,77 | \$1.406,87 | - \$159,01 | \$6.692,48 |
| JPM | \$352,82 | - \$3.378,22 | - \$4.968,51 | \$12.836,16 |

Tabela 5.24: Donosi napovednih modelov na delnicah Russel Top 50 pri začetni investiciji \$10.000 v testnem obdobju

| Delnica | Kupi in zadrži | MA(50,100) | MACD | W2V |
|---------|----------------|--------------|--------------|--------------|
| VZ | \$9.713,13 | - \$458,68 | \$1.250,10 | - \$90,76 |
| T | \$6.461,79 | - \$1.889,04 | - \$6.387,00 | \$1.536,30 |
| UNH | \$44.920,05 | \$18.003,18 | \$17.673,69 | \$43.584,16 |
| AMGN | \$23.283,03 | \$154,68 | - \$308,84 | \$13.356,67 |
| GE | \$5.036,32 | \$2.943,99 | \$9.652,36 | \$3.881,35 |
| CELG | \$27.412,31 | \$6.103,23 | \$4.106,39 | \$17.283,22 |
| CMCSA | \$27.766,76 | \$10.147,41 | \$407,47 | \$24.399,86 |
| KO | \$8.421,94 | \$2.175,98 | - \$3.423,77 | \$1.176,19 |
| MCD | \$11.606,44 | \$1.742,32 | - \$1.463,41 | \$8.612,97 |
| AGN | - \$1.993,17 | - \$1.934,14 | - \$2.338,46 | - \$2.416,12 |
| QCOM | \$5.643,19 | \$5.125,21 | - \$4.300,75 | \$2.084,92 |
| SLB | \$2.353,03 | \$6.444,65 | - \$5.832,95 | - \$2.708,82 |
| HD | \$49.385,92 | \$12.610,17 | \$6.052,85 | \$5,35 |
| BAC | - \$2.805,43 | \$1.590,62 | - \$4.051,77 | \$9.108,43 |
| PFE | \$9.906,60 | \$10.588,31 | \$677,53 | \$9.022,40 |
| WFC | \$5.601,88 | \$4.434,03 | - \$7.759,83 | \$6.839,89 |
| CVX | \$5.060,98 | - \$2.256,97 | \$2.314,22 | \$0,00 |
| UTX | \$11.286,23 | \$7.592,45 | \$5.095,02 | \$3.494,55 |
| MDT | \$3.586,55 | \$3.021,98 | \$2.946,20 | \$1.094,28 |
| HON | \$22.746,48 | \$14.850,36 | - \$583,15 | \$18.197,33 |
| BMJ | \$19.994,72 | \$2.089,88 | \$8.570,57 | \$157,22 |
| BA | \$17.278,83 | \$7.487,06 | \$817,96 | \$4.125,85 |
| IBM | \$7.480,95 | \$4.545,24 | - \$1.584,07 | \$1.424,40 |
| WMT | \$3.290,56 | \$257,12 | - \$3.849,64 | \$3.986,10 |
| AAPL | \$73.940,35 | \$29.638,56 | \$9.705,61 | \$85.059,88 |
| MSFT | \$14.722,34 | \$2.161,22 | - \$59,80 | - \$1.038,90 |
| BRKB | \$8.856,96 | \$8.168,75 | - \$4.624,55 | \$11.791,96 |
| MA | \$20.540,46 | \$8.910,75 | \$608,40 | \$18.084,83 |
| DIS | \$26.464,17 | \$16.746,22 | \$1.539,39 | \$22.774,90 |
| V | \$16.842,45 | \$5.875,98 | \$2.129,09 | \$5.916,25 |
| MMM | \$20.615,30 | \$9.699,65 | \$1.094,79 | - \$452,94 |
| PM | \$1.934,04 | - \$1.617,83 | - \$601,79 | \$1.927,86 |
| INTC | \$12.228,46 | \$3.286,73 | - \$292,18 | \$5.336,48 |
| CSCO | \$6.935,66 | \$5.274,26 | - \$2.023,80 | \$3.484,50 |
| PG | \$3.343,42 | \$172,81 | \$448,23 | \$823,90 |
| GOOGL | \$24.022,95 | \$1.000,41 | \$4.526,77 | \$195,94 |
| UNP | \$21.342,11 | \$32.974,42 | - \$5.868,17 | \$23.017,03 |
| JNJ | \$9.158,11 | \$5.587,77 | - \$814,60 | \$6.156,39 |
| MRK | \$11.607,88 | \$2.739,50 | \$89,90 | \$2.591,13 |
| XOM | \$1.316,88 | - \$776,91 | - \$6.452,51 | - \$2.481,73 |
| MO | \$26.565,02 | \$11.039,3 | \$13.046,51 | \$12.531,12 |
| AMZN | \$121.833,56 | \$32.907,38 | \$2.997,46 | \$81.262,29 |
| ABBV | \$246,24 | - \$2.077,94 | - \$2.462,14 | \$3.657,52 |
| GILD | \$31.654,22 | \$20.098,60 | - \$1.215,54 | \$7.722,35 |
| ORCL | \$14.443,13 | - \$2.395,77 | \$4.578,54 | \$10.926,07 |
| FB | \$8.941,33 | \$5.757,85 | - \$221,55 | \$12.326,12 |
| C | - \$6.715,97 | - \$5.528,77 | - \$9.968,27 | \$9.138,86 |
| CVS | \$21.447,72 | \$7.658,80 | \$1.940,18 | \$17.048,08 |
| PEP | \$6.892,00 | \$714,13 | - \$3.516,71 | \$7.093,98 |
| JPM | \$6.928,75 | - \$1.463,32 | - \$2.505,74 | - \$7.156,60 |

Tabela 5.25: Donosi napovednih modelov na delnicah Russel Top 50 pri začetni investiciji \$10.000 v valiacijskem obdobju

| Metoda | testno obdobje | | validacijsko obdobje | |
|----------------|----------------|------------|----------------------|------------|
| | W | p-vrednost | W | p-vrednost |
| Kupi in zadrži | 2 | <0,0001 | - | - |
| MA(50, 100) | 1 | <0,0001 | 427 | 0,021 |
| MACD | 1 | <0,0001 | 155 | <0,0001 |

Tabela 5.26: Rezultati Wilcoxonovega testa s predznačenimi rangi

Poglavje 6

Sklepne ugotovitve

V našem delu smo se osredotočili na napovedovanje smeri gibanja delnic in trgovanja z le-temi. V uvodnih poglavjih smo predstavili izbrane delnice – Apple, Coca-Cola in Microsoft – ter si pogledali sestavo podatkov za posamezno delnico. Ugotovili smo, da delnice podjetja Microsoft v testnem obdobju padajo, kar se je pokazalo tudi pri primerjalni strategiji *Kupi in zadrži*, ki je v tem obdobju seveda pridelala izgubo.

Preden smo se podali v implementacijo napovednega modela, smo morali definirati trgovalno strategijo, za katero se je izkazalo, da je morda celo bolj pomembna od samega napovedovanja. V poglavju 5.2 smo pokazali, kako lahko z majhno spremembo trgovalne strategije drastično izboljšamo končni rezultat tudi pri relativno slabih napovednih modelih. Predlagana trgovalna strategija temelji na zelo enostavnih predpostavkah, vendar se je tako v fazi testiranja njene kakovosti kot tudi kasneje pri podrobnejši analizi predlaganih napovednih modelov izkazala za zelo dobro. Kljub temu je možnih še veliko izboljšav. V prihodnosti bi želeli v strategijo odločanja o nakupu in prodaji vgraditi tudi funkcijo preprečevanja izgube (angl. *stop loss*), dodatno želimo v okviru odločanja poleg napovednega modela upoštevati tudi že znane in preverjene tehnične indikatorje, kot sta MACD in preprosta tekoča povprečja, kot tudi mnoge druge.

Predlagani napovedni model, ki temelji na obravnavanju *OHLC* podatkov

kot besed, smo razvili v več fazah. V prvi fazi smo želeli praktično neskončno množico *OHLC* podatkov zmanjšati na nabor nekaj 10 različnih besed, kar smo storili s pomočjo japonskih svečnikov in samodejnega gručenja podatkov v skupine le-teh. Pokazali smo, da so si različne gruče med seboj precej podobne, a vseeno dobro določijo različne skupine japonskih svečnikov. Pri zaznavanju japonskih svečnikov je veliko prostora za izboljšave, vsekakor bi bilo treba poleg KMeans gručenja preizkusiti še ostale tehnike, kot so hierarhično gručenje ali DBSCAN.

Pokazali smo tudi, da so v naših učnih in testnih podatkih nekateri japonski svečniki precej bolj pogosti kakor drugi, lastnost, ki jo vidimo tudi pri analizi besedil. V drugi fazi razvoja smo razvili besedni zaklad japonskih svečnikov združili v zaporedje stavkov in te posredovali kot vhodni podatek algoritmu *Word2Vec*. Za omenjeni algoritem smo v uvodnih poglavjih pokazali, da kaže na področju obdelave in razumevanja naravnega jezika veliko potenciala. V zadnji fazi smo opravili strojno učenje in napovedovanje trendov.

Rezultati v poglavju 5.4.5 so pokazali, da predlagani model s pravilno izbiro parametrov v testnem obdobju deluje zelo dobro. Pri vseh treh delnicah smo uspeli poiskati takšne parametre, ki so v testnem obdobju, znotraj simulatorja trgovanja, zaključili obdobje z dobičkom, ki je bil večji od donosov primerjalnih modelov in strategij *Kupi in zadrži*, *Tekoča povprečja* in *MACD*. Parametri, ki so se dobro obnesli v testni množici, so na naše presenečenje delovali dobro tudi v validacijskem obdobju. Ali predlagani napovedni model deluje tudi statistično značilno bolje, smo dodatno preverili poglavju 5.5. Rezultati statističnega preverjanja so potrdili, da s pravilno izbiro parametrov napovedni model deluje statistično značilno bolje od primerjanih modelov. Pri obsežnejšem naboru podatkov (Russell Top 50) se je izkazalo, da znotraj validacijskega obdobja model statistično značilno bolje deluje zgolj v primerjavi z metodama *MACD* ter *Tekoča povprečja*, medtem ko se strategija *Kupi in zadrži* tokrat izkaže tudi za statistično značilno najbolj donosno. Seveda bi lahko rezultate poskušali izboljšati z boljšo trgovalno strategijo oziroma z

vgradnjo zgoraj omenjene funkcije preprečevanja izgube.

Podrobnejša analiza grafov trgovanja in statistična analiza sta pokazali, da predlagani model vsebuje velik potencial za praktično uporabo. Kljub temu pa se ne smemo prenačljiti in zaključiti, da predlagani model zagotavlja dobiček, saj smo s statistično analizo pokazali, da se izbrani parametri v različnih časovnih obdobjih ne obnesejo enako dobro v smislu donosa. Pri analizi grafov smo videli tudi, da učna množica močno vpliva na sam napovedni model, tako kot smo videli na Sliki 5.18, da je napovedni model kljub splošni rasti zelo previden v primerjavi s trgovanjem z delnicami podjetja Apple. Ta problem bi lahko reševali z vključitvijo daljšega učnega obdobja, česar pa zaradi težav dostopnosti podatkov OHLC pred letom 2000 v okviru tega magistrskega dela žal nismo mogli realizirati. Prihodnje izboljšave tako vključujejo tudi uporabo OHLC podatkov drugih delnic v fazi učenja, saj lahko na tak način pridobimo več raznolikih vzorcev.

Pokazali smo, da predlagani napovedni model deluje statistično značilno bolje od primerjanih modelov, vendar pod pogojem, da smo v fazi učenja izbrali primerne parametre. Za konkretne trditve, da metoda dobro deluje tudi v splošnem, je treba opraviti še veliko dodatnih analiz in statističnih testov. Pred tem pa je vsekakor treba drastično povečati učno množico, saj je za dobro delovanje algoritma *Word2Vec* potreben ogromen nabor učnih podatkov [4]. Dodatne izboljšave so možne tudi pri izbiri klasifikacijskega algoritma. Uporabljeni algoritem *Softmax* je precej dobro služil našim potrebam v okviru tega dela, vendar bi boljše klasifikacijsko točnost, manjšo logaritmično izgubo in posledično tudi večje donose lahko pridobili z uporabo naprednejših algoritmov strojnega učenja za klasifikacijo, kot je na primer nevronska mreža, ki jo uporablja tudi algoritem *Word2Vec*.

Ena izmed alternativnih metod napovedovanja, ki bi jo bilo v prihodnje vredno raziskati, bi bila tudi preprosta linearna operacija seštevanja vektorskih predstavitev zadnjih n japonskih svečnikov. Tako bi lahko pridobili napoved jutrišnjega dne oziroma v primeru združevanja zadnjih nekaj tednov tudi tedensko ali mesečno napoved gibanja.

Rezultati predlaganega napovednega modela so pokazali, da je napovedovanje na takšen način ne le možno, temveč prinaša s pravilno izbiro parametrov tudi statistično značilno boljše donose od primerjanih metod. Spodbudni rezultati, ki smo jih pridobili v okviru tega dela, služijo kot motivacija za nadaljnje raziskovanje.

Literatura

- [1] Arman Khadjeh Nassirtoussi, Saeed Reza Aghabozorgi, Ying Wah Teh, and David Chek Ling Ngo. Text mining of news-headlines for FOREX market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment. *Expert Syst. Appl.*, 42(1):306–324, 2015.
- [2] Yifu Huang, Kai Huang, Yang Wang, Hao Zhang, Jihong Guan, and Shuigeng Zhou. Exploiting twitter moods to boost financial trend prediction based on deep network models. In *Intelligent Computing Methodologies – 12th International Conference, ICIC 2016, Lanzhou, China, August 2–5, 2016, Proceedings, Part III*, pages 449–460, 2016.
- [3] Yauheniya Shynkevich, T. M. McGinnity, Sonya A. Coleman, and Ammar Belatreche. Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning. *Decision Support Systems*, 85:74–83, Maj 2016.
- [4] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc., 2013.
- [5] Hércules Antonio do Prado, Edilson Fernalda, Luis C. R. Morais, Alfredo José Barreto Luiz, and Eduardo Matsura. On the effectiveness of candlestick chart analysis for the brazilian stock market. In *17th International*

- Conference in Knowledge Based and Intelligent Information and Engineering Systems, KES 2013, Kitakyushu, Japan, 9-11 September 2013*, pages 1136–1145, 2013.
- [6] Steve Nison. *Japanese Candlestick Charting Techniques: A Contemporary Guide to the Ancient Investment Techniques of the Far East*. New York Institute of Finance, 1991.
- [7] Milad Jasemi, Ali M. Kimiagari, and A. Memariani. A modern neural network model to do stock market timing on the basis of the ancient investment technique of japanese candlestick. *Expert Syst. Appl.*, 38(4):3884–3890, 2011.
- [8] Tsung-Hsun Lu and Yung-Ming Shiu. Tests for Two-Day Candlestick Patterns in the Emerging Equity Market of Taiwan. *Emerging markets finance and trade*, 48(1):41–57, Jan-Feb 2012.
- [9] Takenori Kamo and Cihan H. Dagli. Hybrid approach to the japanese candlestick method for financial forecasting. *Expert Syst. Appl.*, 36(3):5023–5030, 2009.
- [10] Tsung-Hsun Lu. The profitability of candlestick charting in the Taiwan stock market. *Pacific-Basin Finance Journal*, 26:65–78, Jan 2014.
- [11] Karsten Martiny. Unsupervised discovery of significant candlestick patterns for forecasting security price movements. In *KDIR 2012 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, Barcelona, Spain, 4 - 7 October, 2012*, pages 145–150, 2012.
- [12] David Aronson. *Evidence-Based Technical Analysis*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2007.
- [13] David Arthur and Sergei Vassilvitskii. K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM*

Symposium on Discrete Algorithms, SODA '07, pages 1027–1035, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.

- [14] Peter Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, 20(1):53–65, November 1987.
- [15] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.
- [16] Felix Ming Fai Wong, Zhenming Liu, and Mung Chiang. Stock market prediction from WSJ: text mining via sparse matrix factorization. *CoRR*, abs/1406.7330, 2014.
- [17] Jesse Read, Bernhard Pfahringer, Geoffrey Holmes, and Eibe Frank. Classifier chains for multi-label classification. In *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2009, Bled, Slovenia, September 7-11, 2009, Proceedings, Part II*, pages 254–269, 2009.
- [18] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

Dodatek A

Podrobni rezultati modela W2V

| λ | st. zdr. dni | st. svecnikov | sentence_length | logloss | klas. tocnost | Donos simulacija |
|-----------|--------------|---------------|-----------------|---------|---------------|------------------|
| 0.001 | 1 | 70 | 25 | 1.05 | 46% | \$13,078.73 (21) |
| 0.001 | 1 | 70 | 20 | 1.05 | 45% | \$12,881.14 (13) |
| 0.001 | 1 | 70 | 25 | 1.05 | 46% | \$13,078.73 (21) |
| 0.001 | 2 | 50 | 30 | 1.05 | 46% | \$17,732.49 (23) |
| 0.001 | 2 | 70 | 30 | 1.05 | 47% | \$13,163.64 (17) |
| 0.001 | 3 | 50 | 20 | 1.05 | 45% | \$12,133.33 (19) |
| 0.001 | 3 | 70 | 30 | 1.05 | 46% | \$15,949.12 (15) |
| 0.001 | 3 | 100 | 30 | 1.05 | 46% | \$12,363.54 (11) |
| 0.001 | 4 | 30 | 5 | 1.06 | 46% | \$12,404.46 (11) |
| 0.001 | 4 | 40 | 30 | 1.05 | 45% | \$15,074.28 (17) |
| 0.001 | 5 | 30 | 5 | 1.06 | 46% | \$12,478.54 (11) |
| 0.001 | 6 | 30 | 5 | 1.06 | 45% | \$13,671.01 (14) |
| 0.001 | 6 | 70 | 30 | 1.05 | 45% | \$14,014.85 (15) |
| 0.001 | 6 | 90 | 5 | 1.07 | 44% | \$14,838.56 (13) |
| 0.0001 | 1 | 50 | 5 | 1.07 | 44% | \$15,785.35 (17) |
| 0.0001 | 1 | 70 | 10 | 1.06 | 45% | \$12,090.78 (31) |
| 0.0001 | 1 | 80 | 10 | 1.07 | 45% | \$12,810.41 (21) |
| 0.0001 | 2 | 70 | 5 | 1.06 | 46% | \$12,945.73 (40) |
| 0.0001 | 2 | 100 | 20 | 1.05 | 45% | \$13,009.67 (27) |
| 0.0001 | 3 | 40 | 30 | 1.05 | 44% | \$13,981.19 (40) |
| 0.0001 | 3 | 50 | 25 | 1.05 | 46% | \$14,626.90 (18) |
| 0.0001 | 3 | 50 | 20 | 1.05 | 46% | \$15,530.55 (22) |
| 0.0001 | 3 | 50 | 25 | 1.05 | 46% | \$14,626.90 (18) |
| 0.0001 | 4 | 50 | 5 | 1.06 | 45% | \$12,336.05 (26) |
| 0.0001 | 4 | 50 | 10 | 1.05 | 45% | \$13,444.13 (18) |
| 0.0001 | 4 | 50 | 20 | 1.06 | 45% | \$15,689.72 (18) |
| 0.0001 | 4 | 70 | 10 | 1.06 | 43% | \$12,221.11 (30) |
| 0.0001 | 4 | 80 | 10 | 1.06 | 42% | \$12,131.32 (20) |
| 0.0001 | 4 | 90 | 30 | 1.06 | 45% | \$12,417.76 (32) |
| 0.0001 | 4 | 100 | 25 | 1.06 | 45% | \$12,580.91 (22) |
| 0.0001 | 4 | 100 | 25 | 1.06 | 45% | \$12,580.91 (22) |
| 0.0001 | 5 | 40 | 25 | 1.06 | 45% | \$13,146.73 (34) |
| 0.0001 | 5 | 40 | 25 | 1.06 | 45% | \$13,146.73 (34) |
| 0.0001 | 5 | 50 | 5 | 1.06 | 45% | \$13,759.18 (24) |
| 0.0001 | 5 | 50 | 10 | 1.05 | 44% | \$13,177.25 (16) |
| 0.0001 | 5 | 50 | 20 | 1.06 | 44% | \$13,551.83 (18) |
| 0.0001 | 5 | 50 | 30 | 1.06 | 44% | \$13,163.81 (20) |

| λ | st. zdr. dni | st. svecnikov | sentence_length | logloss | klas. tocnost | Donos simulacija |
|-----------|--------------|---------------|-----------------|---------|---------------|------------------|
| 0.0001 | 5 | 70 | 10 | 1.06 | 44% | \$13,890.57 (25) |
| 0.0001 | 5 | 90 | 20 | 1.06 | 45% | \$12,577.07 (26) |
| 0.0001 | 5 | 90 | 30 | 1.06 | 46% | \$14,597.54 (24) |
| 0.0001 | 6 | 40 | 25 | 1.06 | 45% | \$14,256.12 (30) |
| 0.0001 | 6 | 40 | 25 | 1.06 | 45% | \$14,256.12 (30) |
| 0.0001 | 6 | 50 | 5 | 1.06 | 46% | \$17,295.65 (22) |
| 0.0001 | 6 | 50 | 30 | 1.06 | 44% | \$12,664.65 (20) |
| 0.0001 | 6 | 70 | 10 | 1.06 | 44% | \$15,979.03 (21) |
| 0.0001 | 6 | 70 | 25 | 1.06 | 44% | \$13,385.89 (25) |
| 0.0001 | 6 | 80 | 10 | 1.06 | 43% | \$13,121.95 (19) |
| 0.0001 | 6 | 90 | 5 | 1.06 | 44% | \$13,392.07 (36) |
| 0.0001 | 6 | 90 | 25 | 1.06 | 46% | \$12,457.33 (24) |
| 0.0001 | 6 | 90 | 20 | 1.06 | 45% | \$14,889.17 (26) |
| 0.0001 | 6 | 90 | 25 | 1.06 | 46% | \$12,457.33 (24) |
| 0.0001 | 6 | 90 | 30 | 1.06 | 46% | \$12,516.15 (28) |
| 0.0001 | 6 | 100 | 5 | 1.06 | 43% | \$14,745.59 (14) |
| 0.0001 | 6 | 100 | 30 | 1.06 | 43% | \$13,523.75 (29) |
| 0.0001 | 7 | 30 | 10 | 1.06 | 44% | \$12,900.51 (18) |
| 0.0001 | 7 | 30 | 30 | 1.06 | 44% | \$13,821.57 (24) |
| 0.0001 | 7 | 50 | 5 | 1.06 | 45% | \$13,422.72 (20) |
| 0.0001 | 7 | 50 | 10 | 1.06 | 46% | \$17,457.44 (18) |
| 0.0001 | 7 | 50 | 25 | 1.06 | 46% | \$12,093.09 (16) |
| 0.0001 | 7 | 50 | 30 | 1.06 | 44% | \$12,779.58 (18) |
| 0.0001 | 7 | 70 | 10 | 1.06 | 44% | \$15,412.62 (19) |
| 0.0001 | 7 | 90 | 5 | 1.06 | 44% | \$12,428.40 (26) |
| 0.0001 | 7 | 90 | 20 | 1.06 | 43% | \$12,506.03 (22) |
| 0.0001 | 7 | 100 | 5 | 1.06 | 44% | \$13,673.04 (18) |
| 0.0001 | 7 | 100 | 30 | 1.07 | 43% | \$12,478.21 (15) |
| 0.0001 | 8 | 40 | 20 | 1.07 | 44% | \$12,092.59 (17) |
| 0.0001 | 8 | 50 | 5 | 1.06 | 45% | \$16,398.47 (22) |
| 0.0001 | 8 | 50 | 25 | 1.06 | 44% | \$13,922.06 (18) |
| 0.0001 | 8 | 50 | 30 | 1.06 | 44% | \$14,701.60 (18) |
| 0.0001 | 8 | 70 | 10 | 1.06 | 42% | \$14,952.78 (17) |
| 0.0001 | 8 | 70 | 25 | 1.07 | 42% | \$13,594.11 (29) |
| 0.0001 | 8 | 70 | 30 | 1.07 | 42% | \$14,966.66 (19) |
| 0.0001 | 9 | 20 | 30 | 1.06 | 43% | \$12,173.48 (15) |
| 0.0001 | 9 | 50 | 5 | 1.06 | 44% | \$15,379.37 (20) |
| 0.0001 | 9 | 70 | 5 | 1.07 | 41% | \$12,266.73 (29) |
| 0.0001 | 9 | 70 | 10 | 1.07 | 43% | \$15,877.31 (15) |
| 0.0001 | 9 | 70 | 30 | 1.07 | 41% | \$14,078.56 (23) |
| 0.0001 | 9 | 90 | 25 | 1.06 | 43% | \$12,169.55 (19) |
| 0.0001 | 9 | 90 | 20 | 1.06 | 43% | \$13,097.18 (15) |
| 0.0001 | 10 | 30 | 10 | 1.06 | 44% | \$13,419.60 (13) |
| 0.0001 | 10 | 30 | 25 | 1.06 | 44% | \$14,281.39 (11) |
| 0.0001 | 10 | 30 | 30 | 1.06 | 44% | \$14,388.88 (11) |
| 0.0001 | 10 | 70 | 10 | 1.07 | 41% | \$14,633.00 (15) |
| 0.0001 | 10 | 70 | 20 | 1.07 | 41% | \$12,944.20 (17) |
| 0.0001 | 10 | 90 | 5 | 1.07 | 41% | \$12,653.07 (21) |
| 0.0001 | 10 | 90 | 20 | 1.06 | 43% | \$13,449.63 (17) |
| 0.000001 | 2 | 80 | 20 | 1.06 | 44% | \$12,062.96 (41) |
| 0.000001 | 4 | 70 | 5 | 1.09 | 42% | \$12,445.36 (58) |
| 0.000001 | 4 | 70 | 10 | 1.09 | 42% | \$13,989.48 (48) |
| 0.000001 | 4 | 70 | 20 | 1.10 | 41% | \$15,671.77 (52) |
| 0.000001 | 4 | 100 | 10 | 1.10 | 40% | \$12,322.74 (44) |

Tabela A.1: Napovedovanje z Word2Vec metodo z združevanjem - delnice podjetja Apple

| λ | st. zdr. dni | st. svecnikov | sentence_length | logloss | klas. tocnost | Donos simulacija |
|-----------|--------------|---------------|-----------------|---------|---------------|------------------|
| 0.001 | 5 | 90 | 10 | 1.05 | 52% | \$1058.60 (15) |
| 0.001 | 6 | 90 | 20 | 0.97 | 57% | \$1148.54 (11) |
| 0.001 | 7 | 90 | 20 | 0.96 | 60% | \$1143.04 (13) |
| 0.001 | 7 | 90 | 30 | 0.96 | 58% | \$1042.69 (17) |
| 0.001 | 8 | 90 | 20 | 0.96 | 61% | \$1200.11 (11) |
| 0.001 | 9 | 10 | 20 | 1.00 | 59% | \$1006.52 (15) |
| 0.001 | 9 | 20 | 25 | 0.98 | 59% | \$1006.76 (19) |
| 0.001 | 9 | 20 | 25 | 0.98 | 59% | \$1006.76 (19) |
| 0.001 | 10 | 10 | 20 | 0.99 | 60% | \$1050.18 (13) |
| 0.001 | 10 | 30 | 5 | 1.03 | 52% | \$1002.42 (11) |
| 0.001 | 10 | 60 | 25 | 0.97 | 62% | \$1109.57 (13) |
| 0.001 | 10 | 60 | 20 | 0.96 | 61% | \$1069.92 (13) |
| 0.001 | 10 | 60 | 25 | 0.97 | 62% | \$1109.57 (13) |
| 0.001 | 10 | 100 | 20 | 1.00 | 58% | \$1070.58 (13) |
| 0.001 | 20 | 20 | 25 | 0.97 | 59% | \$1158.47 (11) |
| 0.001 | 20 | 20 | 20 | 0.97 | 60% | \$1092.68 (15) |
| 0.001 | 20 | 20 | 25 | 0.97 | 59% | \$1158.47 (11) |
| 0.001 | 20 | 60 | 20 | 0.96 | 60% | \$1129.32 (11) |
| 0.001 | 20 | 60 | 30 | 0.99 | 57% | \$1003.75 (11) |
| 0.001 | 30 | 60 | 5 | 0.96 | 64% | \$1242.70 (13) |
| 0.001 | 30 | 60 | 10 | 0.94 | 63% | \$1230.60 (13) |
| 0.001 | 30 | 60 | 30 | 0.93 | 64% | \$1032.23 (10) |
| 0.001 | 60 | 90 | 5 | 0.97 | 60% | \$1136.52 (12) |
| 0.001 | 60 | 90 | 20 | 0.96 | 59% | \$1095.66 (10) |
| 0.0001 | 5 | 90 | 25 | 0.98 | 56% | \$1009.45 (25) |
| 0.0001 | 5 | 90 | 25 | 0.98 | 56% | \$1009.45 (25) |
| 0.0001 | 8 | 10 | 10 | 1.01 | 56% | \$1003.82 (14) |
| 0.0001 | 8 | 10 | 20 | 1.01 | 56% | \$1008.10 (14) |
| 0.0001 | 10 | 60 | 20 | 0.97 | 57% | \$1053.26 (13) |
| 0.0001 | 20 | 70 | 5 | 1.08 | 45% | \$1026.01 (11) |
| 0.0001 | 30 | 60 | 10 | 0.95 | 58% | \$1027.37 (10) |
| 0.0001 | 30 | 70 | 20 | 0.94 | 57% | \$1109.57 (10) |

Tabela A.2: Napovedovanje z Word2Vec metodo z združevanjem - delnice podjetja Coca-Cola

| λ | st. zdr. dni | st. svecnikov | sentence_length | logloss | klas. tocnost | Donos simulacija |
|-----------|--------------|---------------|-----------------|---------|---------------|------------------|
| 0.001 | 1 | 80 | 10 | 1.07 | 43% | \$1001.12 (12) |
| 0.001 | 3 | 10 | 10 | 1.01 | 49% | \$1055.79 (12) |
| 0.001 | 4 | 30 | 10 | 1.01 | 50% | \$1056.45 (12) |
| 0.001 | 5 | 20 | 10 | 1.02 | 47% | \$1055.13 (16) |
| 0.001 | 5 | 70 | 20 | 1.00 | 49% | \$1047.28 (10) |
| 0.001 | 6 | 100 | 25 | 1.00 | 49% | \$1054.89 (36) |
| 0.001 | 6 | 100 | 25 | 1.00 | 49% | \$1054.89 (36) |
| 0.001 | 8 | 20 | 5 | 1.03 | 48% | \$1001.76 (12) |
| 0.001 | 9 | 60 | 5 | 1.02 | 50% | \$1252.53 (12) |
| 0.001 | 10 | 60 | 5 | 1.01 | 51% | \$1252.79 (10) |
| 0.001 | 20 | 70 | 10 | 0.99 | 51% | \$1054.88 (10) |
| 0.001 | 30 | 90 | 25 | 0.98 | 51% | \$1024.02 (13) |
| 0.001 | 30 | 90 | 25 | 0.98 | 51% | \$1024.02 (13) |
| 0.001 | 30 | 90 | 30 | 0.98 | 52% | \$1231.28 (11) |
| 0.001 | 60 | 80 | 25 | 1.02 | 49% | \$1183.32 (10) |
| 0.001 | 60 | 80 | 25 | 1.02 | 49% | \$1183.32 (10) |
| 0.001 | 60 | 90 | 25 | 1.01 | 49% | \$1018.98 (10) |
| 0.001 | 60 | 90 | 25 | 1.01 | 49% | \$1018.98 (10) |
| 0.001 | 60 | 90 | 30 | 1.01 | 50% | \$1052.10 (10) |
| 0.001 | 90 | 10 | 20 | 1.05 | 47% | \$1010.31 (17) |
| 0.0001 | 7 | 40 | 20 | 1.00 | 52% | \$1011.60 (44) |
| 0.0001 | 8 | 40 | 5 | 1.00 | 51% | \$1065.91 (40) |
| 0.0001 | 8 | 40 | 10 | 1.01 | 51% | \$1269.10 (42) |
| 0.0001 | 8 | 40 | 20 | 1.01 | 50% | \$1153.07 (48) |
| 0.0001 | 9 | 40 | 5 | 1.00 | 51% | \$1170.39 (42) |
| 0.0001 | 9 | 40 | 10 | 1.01 | 51% | \$1187.34 (38) |
| 0.0001 | 9 | 40 | 25 | 1.00 | 52% | \$1273.67 (42) |
| 0.0001 | 9 | 40 | 20 | 1.00 | 51% | \$1357.37 (36) |
| 0.0001 | 9 | 40 | 25 | 1.00 | 52% | \$1273.67 (42) |
| 0.0001 | 10 | 40 | 25 | 1.00 | 53% | \$1195.15 (40) |
| 0.0001 | 10 | 40 | 20 | 1.00 | 53% | \$1340.58 (42) |
| 0.0001 | 10 | 40 | 25 | 1.00 | 53% | \$1195.15 (40) |
| 0.0001 | 60 | 30 | 30 | 1.09 | 48% | \$1131.34 (29) |
| 0.0001 | 60 | 90 | 25 | 1.06 | 50% | \$1162.73 (23) |
| 0.0001 | 60 | 90 | 25 | 1.06 | 50% | \$1162.73 (23) |
| 0.0001 | 90 | 30 | 10 | 1.07 | 48% | \$1025.19 (23) |
| 0.0001 | 90 | 30 | 20 | 1.07 | 48% | \$1119.31 (31) |
| 0.000001 | 10 | 70 | 10 | 1.07 | 50% | \$1026.58 (32) |

Tabela A.3: Napovedovanje z Word2Vec metodo z združevanjem - delnice podjetja Microsoft